SUBIC

Submarine Integrated Control

OFFICE OF NAVAL RESEARCH

GENERAL DYNAMICS CORPORATION
ELECTRIC BOAT DIVISION
GROTON, CONNECTICUT

GENERAL DYNAMICS
Electric Boat Division

EXPERIMENTAL INVESTIGATIONS OF MAN-MACHINE PROCESSING OF INFORMATION VOLUME III

by

Taylor L. Booth
Herbert M. Kaufman
Jerry Lamb
Robert M. Levy
Russell A. Reiss
Howard A. Sholl
Grace Vogelli

University of Connecticut

Examined: _____
J.W. Herring
SNBIC Project Manager

Approved: _____
Dr. A.J. van Woerkom
Chief Scientist

U417-68-098
October 1, 1968

ABSTRACT

The aim of this project is to provide basic knowledge of the methods which may be used by a man-computer system to detect the presence of a target, using data from a passive sonar receiver. This research consists of analytical studies to evaluate important system parameters and experimental investigations measuring operator performance under various operating conditions.

The first two reports in this volume describe the effects of pattern variations on human pattern recognition. The results measured the operator's ability to visually detect patterns differing in shape and to detect patterns generated by statistically dependent sequences.

The second two reports deal with basic human information processing and describe the testing of a predictive model for reaction time to visual stimuli and a test of the effects of number of stimuli on memory span.

GENERAL TABLE OF CONTENTS

FOREWORD

The work described in this report was accomplished by members of the Department of Electrical Engineering, University of Connecticut, under subcontract to the SUBIC Program (contract NOnr 2512(00)) during the period from July 1967 to July 1968. The Office of Naval Research is the sponsor and General Dynamics Electric Boat division is the prime contractor. LCDR E.W. Lull, USN, is Project Officer for ONR; J. W. Herring is Project Manager for Electric Boat under the direction of Dr. A. J. van Woerkom, Chief Scientist of the Applied Sciences Department.

INTRODUCTION

The goal of the General Dynamics Electric Boat division research project at the University of Connecticut is to provide basic knowledge concerning the methods which may be used by a man-computer system employing data from passive sonar receivers to detect the presence of a target. This research consists of analytical studies to evaluate important system parameters and experimental investigations measuring operator performance under various operating conditions.

The reports in this volume are divided into two groups; the first deals with pattern detection on a cathode ray tube display, while the second group is concerned with visual information processing.

The first report, No. 23, describes an experiment in which target shape (line, rectangle, or square), target orientation (horizontal and vertical) and signal-to-noise ratio (three levels) were varied. Time to decide if a target was present was dependent on signal to noise and target shape; operator "noise" was independent of all parameters.

Report No. 24 describes three experiments on pattern recognition with dependent statistical sequences. Several findings are reported, generally showing that operator noise and detection performance are poorer then for equivalent independent sequences.

Report No. 25, the first report in the second group, describes a model to predict reaction time from individual stimulus information and an experiment run to validate the model. The results supported the model; previous experimental results were also analyzed by the model.

Report No. 26 describes an experiment to test the effect of a number of different possible symbols to be recalled and information load per symbol on immediate memory. The results generally showed that a constant number of symbols was recalled regardless of the number of possible different symbols. The one condition which did not show this result is examined in light of a coding scheme that subjects could use.

GB/EB-23

U N I V E R S I T Y O F C O N N E C T I C U T

S T O R R S, C O N N E C T I C U T

GENERAL DYNAMICS/ELECTRIC BOAT RESEARCH PROJECT

EFFECT OF PATTERN SHAPE AND ORIENTATION IN VISUAL PATTERN DETECTION

BY

Howard A. Sholl

Department of Electrical Engineering

April, 1968

Acknowledgement is due the Office of Naval Research which supported this program through a prime contract (NOnr 2512(00)) with General Dynamics/ Electric Boat as a part of the SUBIC (Submarine Integrated Control) program.

EFFECT OF PATTERN SHAPE AND ORIENTATION IN VISUAL PATTERN DETECTION

ABSTRACT

This report is concerned with an experimental investigation of the effects of pattern characteristics on man's ability to detect visual signals in noise. Subjects were presented a two-dimensional random dot display and asked to indicate the presence or absence of a signal. Target shapes presented were lines, rectangles, and squares, both vertically and horizontally oriented, and at three signal to noise ratios. The standard deviation of the decision uncertainty - operator noise - was found to be essentially independent of target shape, orientation, and signal to noise ratio. Decision time was independent of orientation, but varied with both shape and signal to noise ratio.

Effect of Pattern Shape and Orientation in Visual Pattern Detection

1.0 Introduction

The general problem being considered here is the development of a method of determining how well a man can detect a visual pattern in a noisy environment. The solution to the problem must be a two step process: first, determining from what pattern characteristics the subject extracts information to guide his detection decision; and second, determining how these information-carrying characteristics interact to produce a final decision. This paper is concerned with the first of these problems.

Previous research in this area has occurred in both physiological and psychological studies. In general there are a few areas of correlation. First of all, it has been found that in the visual cortex of animals such as the rabbit and cat the architecture of the ganglion cells is such that some individual cell structures respond to line stimulation only at specific orientations.[1,2] It is not known whether or not there is an overabundance of these cells optimized at any orientation, but a logical assumption is that they may be distributed in such a manner as to allow such animals to see, equally-well, lines of any orientation. The implication here is that the human's visual system may be constructed in a similar manner. It has been known for some time that visual orientation significantly affects the recognition ability of people.[3] It has been shown that people more readily recognize vertically-oriented patterns than horizontally-oriented patterns.[4] Thus far the explanation of this phenomenon has consisted of the theory that people do not recognize patterns as readily when they are presented out of their normal context. Since, in general, most real life patterns are structured somewhat symmetrically about the vertical axis, it may be true that people are not more capable of recognizing vertically-oriented

patterns, but just more accustomed in doing so. There is some evidence that this may be true. Henle[5] found that an initial difference in recognizing ability between differently oriented patterns disappeared with further training. As a whole, the determination of the effect of orientation on the ability of people to recognize patterns has not clearly been explained. The question that is being raised is, "Are people more capable of seeing vertically or horizontally oriented patterns, perhaps because of the basic cellular structure of the visual system?"

Other visual pattern characteristics which may influence a person's detection decision are contrast between bordering areas, and the shape of the pattern presented. A line can be considered as the edge between two contrasting areas, and the line intensity can be measured as the amount of contrast present. Pattern shape, although a somewhat vague area to define, is included in this investigation to compare man's detection capability of lines with that of areas containing the same information content.

The remainder of this paper investigates these areas---pattern shape, intensity, and orientation---by comparing man's ability to detect visual patterns from a noisy environment with that of an ideal detector.

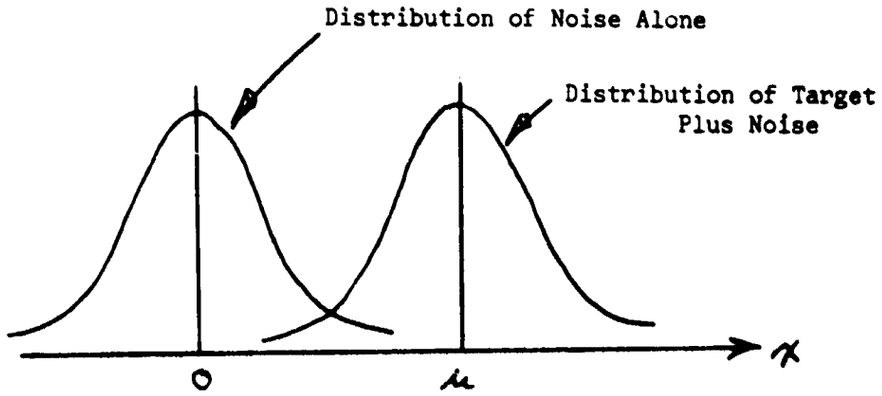
2.0 Description of Display System

The equipment used in the experiment consisted of a cathode-ray tube, random-dot display controlled by a PDP-5 computer. A detailed description of the display system is given in reference 6. The displays used in this experiment were two-dimensional random-dot patterns (72 rows x 72 columns) in which 72 cells were assigned as a target ---line, rectangle, or square. The background noise was controlled by sampling a Gaussian noise source about the mean to determine whether or not a specified cell should be intensified. Sampling the same noise source at a different level determined whether or not a target cell should be intensified. A push button matrix was available for subject responses to the presented displays, and the computer was used to store and process data as the experiment progressed.

3.0 Ideal Detector

It is quite desirable in any research effort to determine a basis for performance which can be used as a measure of the quality of the outcome of an experiment or study. One possibility is to obtain a large amount of previous information in the area of interest and use this as a basis of comparison. Another approach is to determine the ideal results of an experiment and find out how the actual results compare with the ideal. In general, the latter method is to be preferred, because specific areas which may be lacking are more apt to be evident and because the latter method more readily lends itself to modelling.

An ideal detector can be defined as a device which counts the number of intensified cells in the target area and compares this with a predetermined optimum threshold to form a target, no target decision. The decision amounts to deciding whether or not the target plus noise, or just noise alone is present. A detailed treatment of this decision process can be found in reference 7.



Signal-to-noise ratio = $\frac{u}{\sigma}$ where σ = standard deviation of the distribution

Figure I Noise and Target Distributions

The state of each cell in the target area is determined by sampling either of the above Gaussion Distributions about the mean of the noise.

If the noise alone is present:

P_0 (prob. of an intensified point) = $Q_0 = 0.5$

If the target plus noise is present:

$$P_1 = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-(1/2)\left(\frac{x-u}{\sigma}\right)^2} \, dx$$

$$Q_1 = 1-P_1$$

Using 72 samples -- the entire target area---, and applying optimum decision theory to form the likelihood ratio:

$$L = \frac{P_1^{N_1}(Q_1)^{N-N_1}}{P_0^{N_1} Q_0^{N-N_1}}$$

where N = 72, total number of target area cells
N_1 = the number of intensified cells

The decision task is now:

$L \geq Lth$ decide target

$L < Lth$ decide no target

where the Lth is a threshold

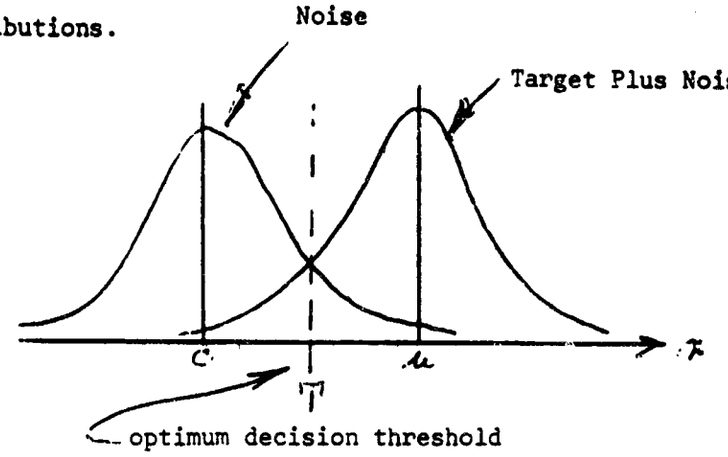
Using the Bayes criteria for equal costs and an <u>a-priori</u> probability of 0.5, the optimal Lth = 1.

Solving for N_1,- the decision threshold;

$$N_1 = \frac{-N \log 2Q_1}{\log P_1/Q_1}$$

The expected results for the ideal detector in a decision task can now be determined by calculating, based on the optimum decision threshold, the detection,

false alarm, correct dismissal, and false dismissal probabilities defined below where the discrete binomial distributions are approximated by uniform Gaussian distributions.



detection probability (decision target when target is present)

$$D = \frac{1}{\sqrt{2\pi}} \int_{T}^{\infty} e^{-(1/2)(\frac{x-u}{\sigma})^2} dx$$

false alarm probability (decision target when no target is present)

$$F = \frac{1}{\sqrt{2\pi}} \int_{T}^{\infty} e^{-(1/2)(\frac{x}{\sigma})^2} dx$$

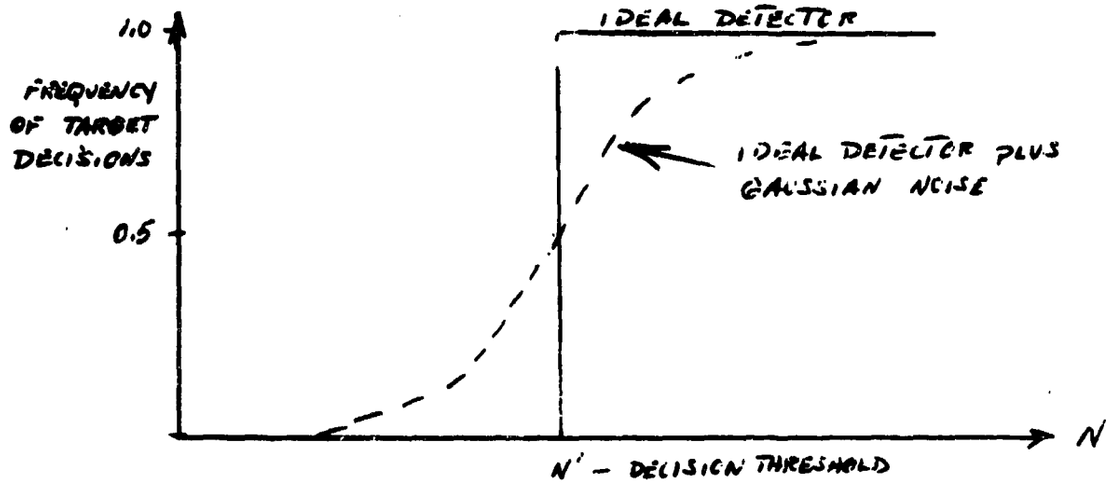
correct dismissal (decision no target when no target is present)

$$CD = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{T} e^{-(1/2)(\frac{x}{\sigma})^2} dx$$

false dismissal (decision no target when a target is present)

$$FD = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{T} e^{-(1/2)(\frac{x-u}{\sigma})^2} dx$$

The decision process of a person can be likened to that of an ideal detector to which a Gaussian noise source has been added.[7]



$N \sim$ (# of intensified points in the target area)

Fig. II Nonideal Decision Process

The ability of a person to make a decision can be measured by determining the standard deviation of this operator noise under different conditions.

4.0 Experimental Design -

The pattern characteristics selected for the experiment were the following:

Shape: line,rectangle, square

Orientation: vertical, horizontal

Intensity: -3.5db, -8.5db, -14.5db signal to noise ratios

Three shapes were included to provide an intermediate target area between a line (minimum area) and a square (maximum area). Oblique orientations were avoided because of the difficulty in obtaining equal oblique dot spacing in a rectangular dot matrix. The signal-to-noise ratios were selected to take advantage of past experimental data for comparison purposes. Each possible combination of these factors was used as the basis of an experimental session. Four subjects were used in an alerted operator, no feedback, signal detection task in which, during each of the fifteen experimental sessions, one hundred displays were randomly presented (fifty target, fifty no-target). The subjects' task was to decide whether or not a target was present. Initially before each session, the subject was given a brief training run to affix his decision threshold at or near that of the ideal detector. The training session consisted of ten patterns with feedback which allowed the subject to reexamine the pattern after learning the outcome of his decision. If the subject desired, the training session was repeated.

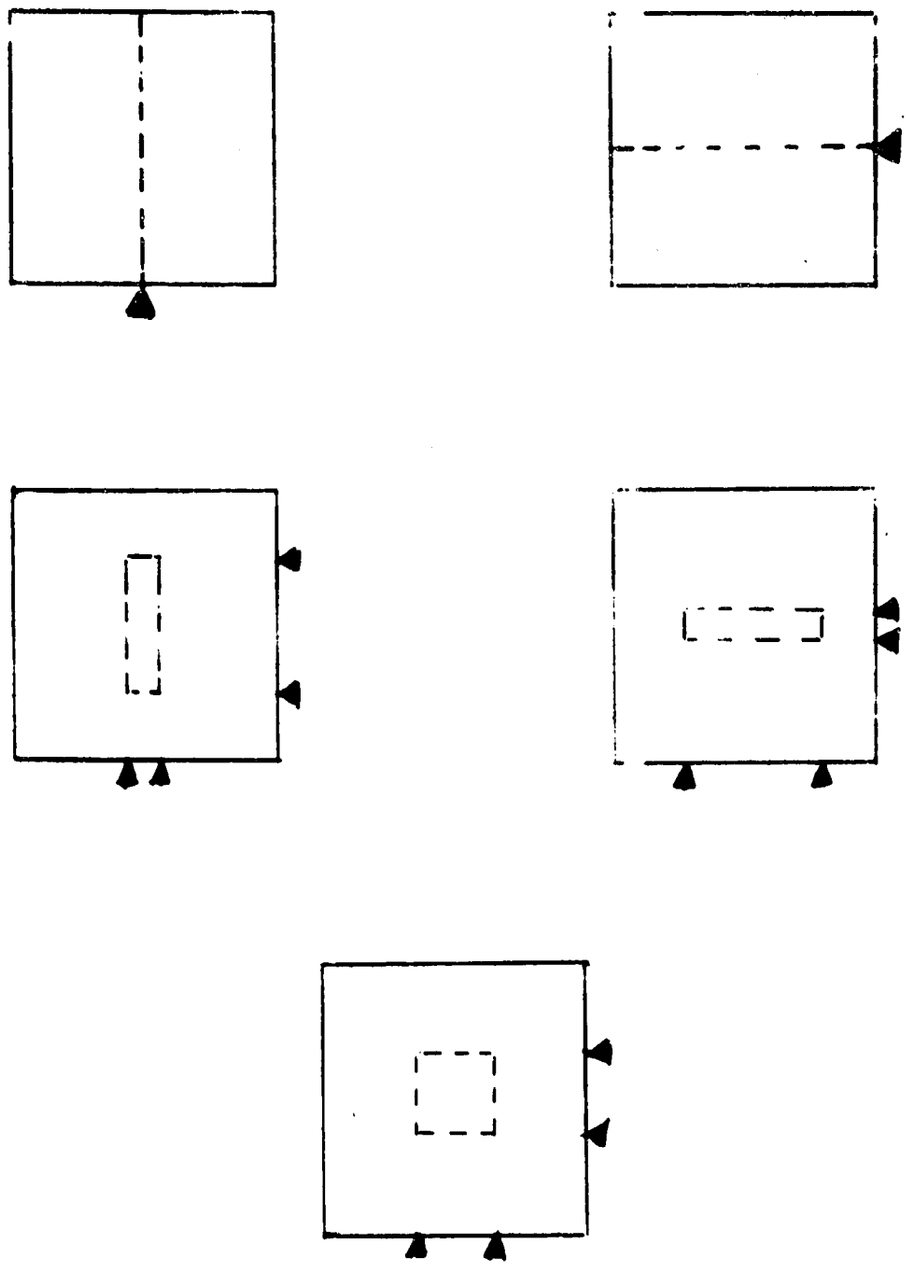


Fig. III TARGET PATTERNS

The target areas were indicated by markers along the bottom and right hand side of the display. (see Fig. III). The overall matrix size and intensity were preset before each session. The subjects were told not to waste time trying to locate the exact target perimeter, but rather to scan the target area denoted by the markers and then indicate their decision by depressing one of two buttons. The data collected consisted of the detection time, to the nearest tenth of a second, and both the ideal detector and operator decisions for each target display. After a session was completed, the computer printed out the experimental results with the following format:

number of intensified points in the target col.	target area source distribution (T or N)	Ideal detection decision (T or N)	operator decision (T or N)	decision time

(for all displays)

	ideal det	operator
detection probability D		
false alarm probability F		
correct dismissal CD		
false dismissal FD		

5.0 Results - Conclusions

The data analysis plan was to determine the standard deviation of the distributions representing the subject's decision characteristics, as a measure of his detection ability. In order to have a large number of data samples and obtain results typical of an average subject, it was desirable to pool all subject's data for each condition. However, initial data analysis of the individual subject's decision characteristics revealed that in spite of the attempt to reduce the between subject decision threshold variation by initial training a significant difference persisted. Thus any attempt to pool the data must first take this effect into consideration by subtracting from each set of data the mean of its assumed-Gaussian distribution. This was accomplished by writing a Fortran program which will find the one Gaussian approximation which best fits the data points using a minimum mean square error criterion. (see Appendix I) Now the means of the individual distributions could be determined and subtracted, and the data pooled for an investigation of the decision uncertainty--operator noise--characteristics. The results are shown below in Table I.

In general the results show that the subjects could detect a target imbedded in noise almost equally well over the range of parameters considered. Effects of orientation are negligible, and only a slight difference in avg. decision uncertainty was evident over the signal-to noise ratio range. The most difficult shape appeared to be rectangular; the easiest a line. However,the manner in which the patterns were presented may have contributed to this result. The location of the pattern in the matrix was indicated to the subject by markers along the bottom and right hand side of the display, (see FigureIII). For lines, all the points in the target area were easily locatable by the subject by scanning along the identified line. For rectangular and square targets, the

Signal/Noise Ratio

	-3.5 DB	-8.5 DB	-14.5 DB	Avg.
Vertical Line	4.8	4.4	3.7	4.30
Horizontal Line	4.4	4.4	4.5	4.43
Vertical Rectangle	4.5	4.4	5.8	4.90
Horizontal Rectangle	4.5	4.8	4.9	4.73
Square	4.4	5.0	4.5	4.63
Avg.	4.52	4.60	4.68	4.60

Orientation:	Vertical	4.60
	Horizontal	4.58
Shape:	Line	4.37
	Rectangle	4.82
	Square	4.63

Each Entry Represents The Standard Deviation of The Subjects' Decision Uncertainty In The Number of Points In The Target Area.

TABLE I POOLED OPERATOR NOISE RESULTS

subjects had the problem of identifying all four edges of the target area. Since they were told not to attempt to accurately locate the target perimeter but rather just scan the indicated area, they had the additional uncertainty of exactly what points were considered the target. Thus, in general, they could be expected to either use a smaller sample for the decision, or perhaps include some points outside of the target area. On this basis, the small difference between the results for different shapes does not seem significant.

Decision time was considered by determining first the average over all subjects for each condition. (see Table II), and then determining the average variation in decision time as a function of the number of points in the target area for each condition. (see figure IV)

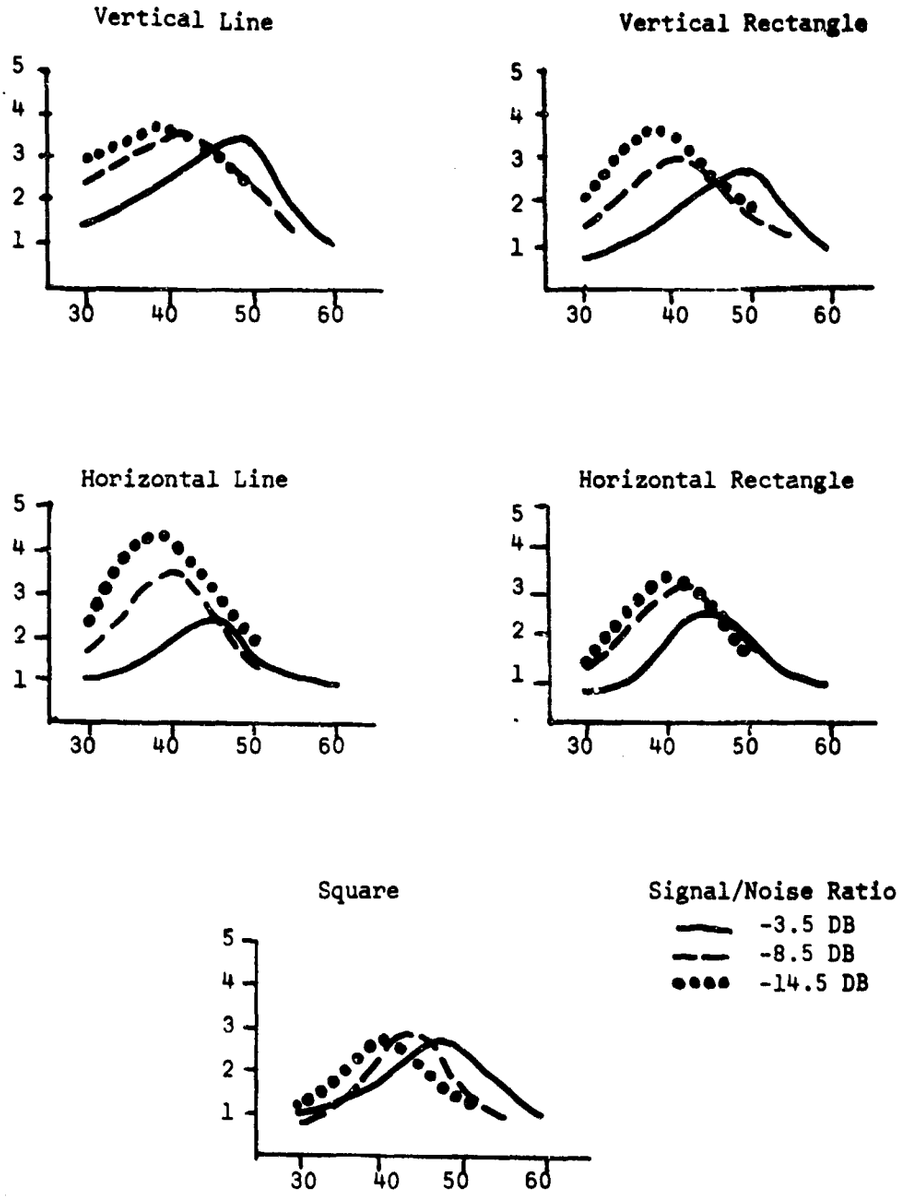


FIGURE IV DECISION TIME VS NUMBER OF TARGET AREA POINTS

Signal/Noise Ratio

	-3.5 DB	-8.5 DB	-14.5 DB	Avg.
Vertical Line	2.32	3.00	3.42	2.91
Horizontal Line	1.63	2.67	3.93	2.74
Vertical Rectangle	1.75	2.62	3.06	2.48
Horizontal Rectangle	1.44	2.35	2.82	2.20
Square	1.42	1.88	2.17	1.82
Avg.	1.71	2.50	3.08	2.43

Orientation:	Vertical	2.69
	Horizontal	2.47
Shape:	Line	2.82
	Rectangle	2.34
	Square	1.82

TABLE II AVERAGE DECISION TIME (SEC.)

From the preceding data several effects are apparent. As should be expected, average decision time increased as the signal/noise ratio decreased, and asymptotically approached a constant. The slight difference between vertical and horizontal conditions is not significant since the individual subject's data does not consistently show the same result. However, as the shape of the target changed in the direction of decreasing perimeter (line, rectangle, square), the decision time decreased significantly. There are two possible reasons why this might be true: first, there may be a difference in the time required to scan the expected target area before the decision is made; and second, if the scanning times are not different, the subject must be processing the information in a different manner. It is evident, from Figure IV , that for obvious decisions -- ones with extremely low or high target point density -- that the decision times do not significantly differ, implying that the scanning times do not significantly affect the decision time. The behavorial

explanation of this result may be the following: When observing a line target the subject must base his decision on two partitions of information: the number of target points presently in view; the recall from memory of the previously scanned target points. When observing a more compact target shape, a larger portion of the target is within view and less memory recall is necessary. On this basis the shorter decision times for more compact target patterns indicated that more rapid decisions may be made when less memory processing is required of the operator.

In conclusion, it appears that

1. Orientation (vertical vs horizontal) has little effect on a subject's decision.
2. Target shape does not affect a subject's ability to make a correct decision, but may alter the manner in which he processes target information. In general the decision time decreases when the target information is presented in a more compact shape.
3. Signal/noise ratio (contrast), over the range considered, does not significantly affect a subject's ability to make a consistent decision, but lengthens the decision time as the target strength decreases.

Appendix I

Best Fit Gaussian Approximation
PDP-5 Fortran

Program Description

This program was written to simplify and improve the curve fitting problem of approximating a psychometric function with a Gaussian Distribution. Essentially, the program begins with an estimate of the mean and standard deviation of a set of data, and iteratively varies the mean and standard deviation, in that order, until a mean square error measure is minimized. It was found, experimentally, that for the resolution of the program (0.1), three iterations were sufficient. Total running time is about 3-5 minutes. Output results are printed on the ASR-33.

Operating Instructions:

1) Load Rim loader at 0020*1

2) Load Dec 8-2U binary loader via Rim

3) Load Fortran operating system

4) Change location $(0404)_8$ to $(7000)_8$

5) SA = 200 Press load address

6) Turn ASR-33 on line, punch off.

7) Insert INTerpretive BFGA program in High Speed reader

8) Enter 2000 in the switch register

9) Press start - program will load and halt with AC=0.

10) Press Continue and load data*2.

*1 - see Rim loader

*2 - see Data format

An output of "mean square error = " will occur for each iteration. If the error is the same for two successive type outs the program has converged on the best solution. If the error has not repeated itself at the program completion, reenter the data in the same format but use the "new" estimates (results of the first run).

Operating Description:

Once the program has been loaded, operation will commence as soon as sufficient data has been introduced. Data may be initially on paper tape typed in ASC-11 form in the proper format*2 or it may be entered from the keyboard as the program is running. If an error is made during input data:

1) press RUB OUT and the program will ignore the preceding word

or, 2) stop the computer and restart at SA=0201 - then reenter the complete data.

Numbers are separated by commas or carriage returns.

This program is in a continual loop so that when a set of data has been processed, a new set may be immediately entered.

	Rim
0020	6014
0021	6011
0022	5021
0023	7300
0024	6012
0025	7106
0026	7006
0027	7510
0030	5020
0031	7006
0032	6014
0033	6011
0034	5033
0035	6012
0036	7420
0037	3442
0040	3042
0041	5020

Data Format

- order -

1	Code number (any number
2	number of data points
3	Est. of mean
4	Est. of Std. Dev.
5	Variable, rate of occurrence - (one data point)
	" , " "

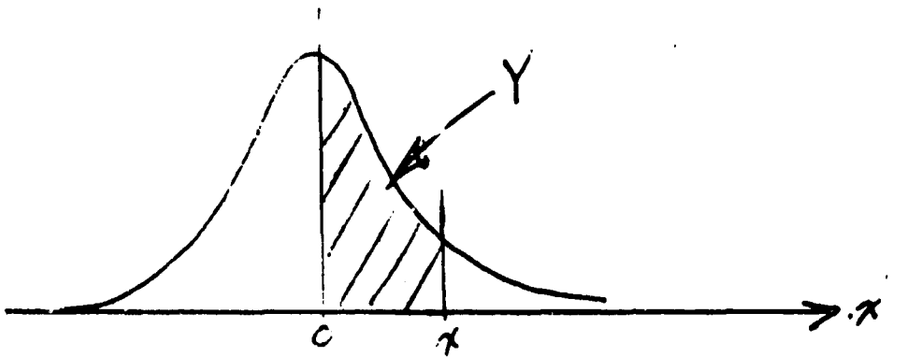
Example of data input format and output results,

```
0111,11,39,4
34.5,0,36.5,.133,38,.17,39,.33,40,.4,41,.375,42,.625,43,.7,44,.8,46,.86
48.5,1
Mean square error = +0.257997E-1
Mean square error = +0.25799E-1
Mean square error = +0.257997E-1
+111
STD DEV = +0.389999E+1
Mean = +0.411997E+2
```

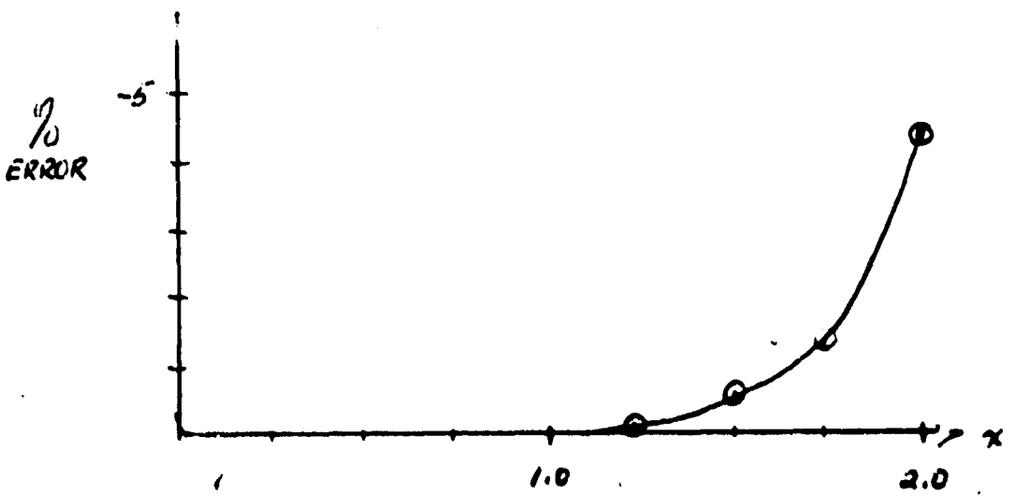
Gaussian Approximation -

The area under the normalized Gaussian curve is calculated by the following polynomial

$$Y = .398\,x - .0663x^3 + .00995x^5 - .00118x^7$$



Approximation Error

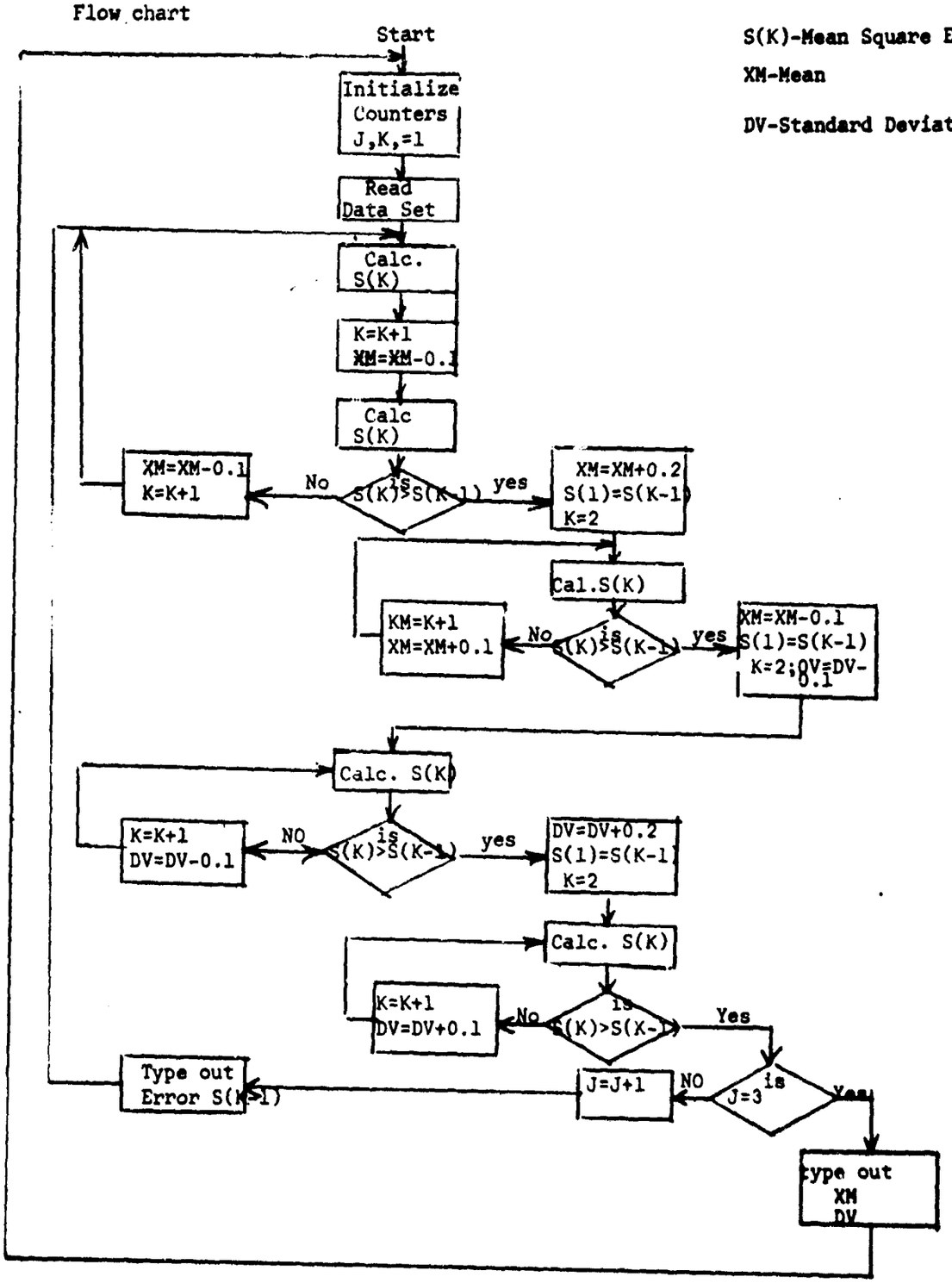


Flow chart

S(K)-Mean Square Error

XM-Mean

DV-Standard Deviation



FORTRAN Listing

```
C     ; Best Fit Gaussian Approximation
      ; Dimension X(20), P(20), C(8), S(70), D(20), Y(20),
      ; C(1)=-.118E-02
      ; C(2)=0
      ; C(3)= .995E-02
      ; C(4)=0
      ; C(5)= -.663E-01
      ; C(6)=0
      ; C(7)= .398
      ; C(8)=0
26    ; Accept 3, T, N, XM, DV
3     ; Format (E, I, E, E)
      ; DO 4 I=1, N
      ; Accept 5, X(I), P(I)
5     ; Format (E, E)
4     ; Continue
      ; DO 29 J=1,4
      ; K=1
      ; M=1
6     ; SM=0.
      ; DO 7 I=1,N
      ; Y(I)=(X(I)-XM)/DV
      ; D(1)=Y(I)*C(1)+C(2)
      ; DO 8 L=3,8
      ; D(L-1)=Y(I)*D(L-2)+C(L)
8     ; Continue
      ; AR=0.5+D(L-1)
      ; IF(AR-P(I)) 27,28,28
27    ; SM=SM+(P(I)-AR)**2
      ; Go to 7
28    ; SM=SM+(AR-P(I))**2
7     ; Continue
      ; S(K)=SM
      ; Go To (9,10,11,12,13),M
14    ; M=M-1
9     ; K=K+1
      ; XM=XM-0.1
      ; M=M+1
      : Go to 6
10    ; If (S(K)-S(K-1)) 14,14,15
15    ; XM=XM+0.1
      ; S(1)=S(K-1)
      ; K=1
18    ; K=K+1
      ; M=M+1
      ;XM=XM+0.1
      ; Go To 6
11    ; If (S(K)-S(K-1)) 16,16,17
16    ; M=M-1
      ; Go To 18
17    ; XM=XM-0.1
      ; S(1)=S(K-1)
      ; K=1
19    ; K=K+1
```

```
     ;  M=M+1
     ; DV=DV-0.1
     ; Go To 6
12   ; IF (S(K) - S(K-1)) 20,20,21
20   ; M=M-1
     ; GO TO 19
21   ; DV=DV+0.1
     ; S(1) = S(K-1)
     ; K=1
24   ; K=K+1
     ; M=M+1
     ; DV=DV+0.1
     ; GO TO 6
13   ; IF (S(K) - S(K-1)) 22,22,23
22   ; M=M-1
     ; GO TO 24
23   ; DV=DV-0.1
     ; Type 1, S(K-1)
1    ; Format (/"Mean Square Error=",E)
29   ; Continue
     ; Type 25, T, DV, XM
25   ; Format (I,/,"STD DEV=", E,/,"Mean=",E)
     ; GO TO 26
     ; STOP
     ;END
```

References

1. Hubel, D. H., and Wiesel, T. N.

   "Receptive Fields, Binocular Interaction, and Functional Architecture in the Cat's Visual Cortex"
   Jour. of Physiology, 1962, vol. 160, pp 106-123.

2. Levick, W. R.

   "Receptive Fields and Trigger Features of Ganglian Cells in the Visual Streak of the Rabbit's Retina"
   Journal of Physiology, 1967, vol. 188, pp 285-307.

3. Dearborn, G. V. N.

   "Recognition Under Objective Reversal"
   Phychol. Rev., 1899, 6, 395-406.

4. Fitts, Mezer, Rappaport, Anderson, Leonard.

   "Stimulus Correlates of Visual Pattern Recognition; a Probability Approach"
   J. Exp. Psychol., 1956, 51, 1-11.

5. Henle, M.

   "An Experimental Investigation of Past Experience as a Determinant of Visual Form Perception"
   J. Exp. Psychol., 1942, 30, 1.22.

6. Brazeal, E. H., Jr. and Booth T. L.

   "A High-Speed Man-Computer Communication System"
   presented at Spring DECUS meeting, April, 1967.

7. Brazeal, E. H., Booth, T. L.

   "Operator Noise in a Discrete Signal Detection Task"
   IEEE, Tran. on Human Factors in Electronics, vol. HFE-7, no. 4, pp 164-173, Dec. 1966.

GD/EB-24

UNIVERSITY OF CONNECTICUT

STORRS, CONNECTICUT

GENERAL DYNAMICS/ELECTRIC BOAT RESEARCH PROJECT

THE EFFECT OF INTER-SYMBOL DEPENDENCIES ON VISUAL PATTERN DETECTION

by

Russell A. Reiss

Taylor L. Booth

Department of Electrical Engineering

June 1968

Acknowledgement is due the Office of Naval Research which in part supported this research through a prime contract (NOnr 2512(00)) with General Dynamics/ Electric Boat as a part of the SUBIC (Submarine Integrated Control) program.

Preface

When analyzing or designing a man-machine system used to perform signal detection or pattern recognition, it is important not only to know the specifications of the computer and other hardware, but also the capabilities and limitations of the human operator. Visual displays generated by statistical processes provide one means of controlling the information presented to the operator, and thereby studying his performance. While other workers have used visual displays generated only by statistically independent processes, this thesis studies the effects of inter-symbol dependencies on human visual information processing ability. In particular, the range of human sensitivity to dependent information, the form of operator noise, as compared to an ideal detector, and the relative utility of statistically independent and dependent information are determined. Also, a method of generating Markov sequences by a small scale digital computer is discussed.

This work was supported in part by an NDEA (National Defense Education Act) Title IV fellowship and also the Office of Naval Research through a prime contract (NOnr 2512(00)) with General Dynamics/Electric Boat as a part of the SUBIC (Submarine Integrated Control) program.

Table of Contents

Table of Contents (Continued) Page

LIST OF TABLES

List of Figures

Chapter 1
Introduction

1.1 The Problem Under Investigation

An accurate description of human visual information processing capabilities, and knowledge of the factors affecting human performance in visual detection tasks, are particularly important if the human is to be successfully integrated with a computer in a man-machine signal detection or pattern recognition system.

In an attempt to analyze the human as a visual information processor, investigators have used displays similar to Figure 1.1, composed of an array of dots generated by statistical processes, in order to control the information in the stimuli. In the typical "alerted operator" detection task all columns, except one near the center, called the target column, represent a random background. The "target", if it is present, appears in the marked column (target column) as a difference in some statistical parameters, for example, the number of intensified points. The operator's task is to determine the presence or absence of a target, or to classify the target column on the basis of some subjective measure. In general this work has been limited by the basic assumption that successive points in the display are statistically independent. In this thesis human visual detection performance is analyzed using patterns generated by dependent statistical processes in order to determine the human's ability to use information provided by inter-symbol dependencies. The three general areas investigated are:

1. The range of human sensitivity to visual dependent information.
2. The form of human "operator noise" in a visual detection task

* note: numbers in parentheses refer to references listed in the bibliography.

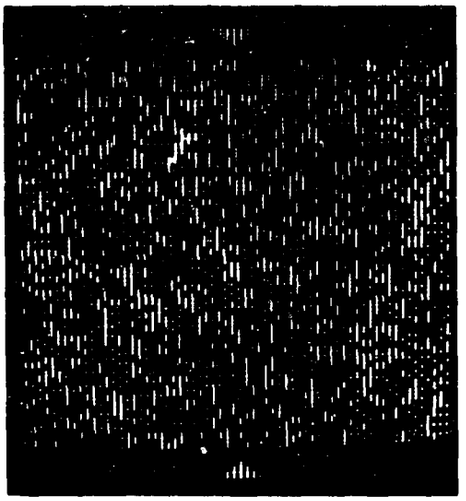


Figure 1.1

Typical Statistical Display Used in Human Visual Information Processing Experiments

with dependent information.

3. The relative utility of independent and dependent visual information to the human operator.

1.2 Background

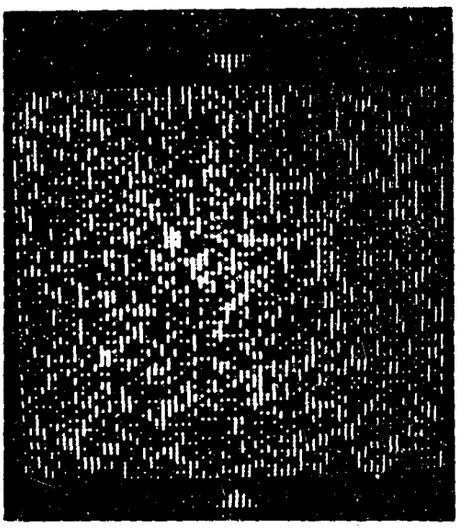
Recent work by Kaufman, Levy, Booth, and Glorioso, (1)*, has considered many aspects of the problem of integrating a small scale digital computer and a human operator to combine the high speed processing and display control capabilities of the computer with the visual detection capabilities of the human operator. The basic display used in these investigations was an array of binary dots on the face of a cathode ray tube as shown in Figure 1.1. All columns except the target column were essentially generated by a statistically independent binary process with P(0)=P(1)= 1/2. The target column was generated by this same process for the no-target condition, and was obtained by increasing P(1) under the condition of target present. The basic assumption of these workers was the statistical independence of each point in the display. For such patterns it may be easily shown (2) that an optimum detector need only count the number of intensified points in the target column and compare this number to a threshold determined by the statistics of the underlying processes, the a priori probabilities of the occurrence of target and no target conditions, the costs associated with each decision, and the desired detection probability. It is not necessary for the optimum detector (in this case)to consider higher-order statistics arising from inter-symbol dependencies in the pattern.

Brazeal and Booth (2), in 1966, considered the problem of "operator noise" in an alerted operator signal detection task. They found that the operator could be modeled as an "optimum detector" with an added noise

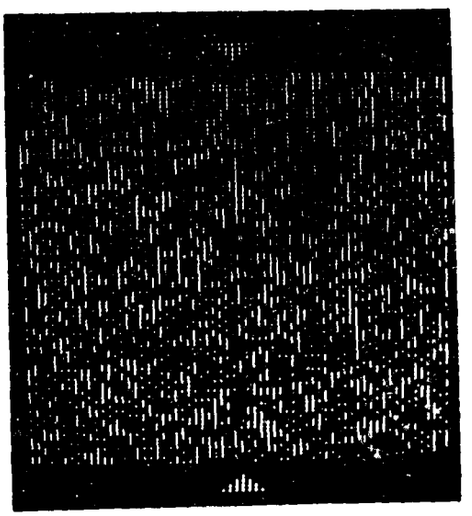
source. The operator noise was found to be Gaussian (normally) distributed with a mean which tended to zero with sufficient training.

This work was extended by Moran (3) to curved targets, while Glorioso (4) developed a stochastic model which describes the dynamics of the human operator and his ability to learn, adjust decision thresholds, etc.

In general, information contained in the first-order statistics of a display is only one component of the total visual information. In addition to this component, higher-order information may be present when there exist dependencies between the symbols. In this paper the term "higher-order statistics" is used to mean probability distributions of sequences of symbols of length greater than one. First-order statistics refer to sequences of length one, the individual symbol frequencies, second-order statistics refer to sequences of length two, and so forth. Consider the two displays shown in Figure 1.2. Each of these displays has 84 rows and 64 columns of binary dots, with the target column marked by arrows. Each target has exactly 42 intensified points ("1's), which is the expected number of intensified points in the other 63 "noise only" columns. The target column of Figure 1.2b, however, has eight more sequences of two consecutive intensified points ("11's") than the target column of Figure 1.2a, which has a total of 21 "11" sequences. An "optimum" first-order detector, making use of only first-order statistics, would view these two target columns as exactly the same since they both have exactly 42 "1's". Although an untrained observer may not be able to distinguish between these two target columns, it is a simple matter for an operator who is trained to look for cues such as clusterings of 1's and 0's to use this dependent information to distinguish between the two



a) N_1=42, N_{11}=21



b) N_1=42, N_{11}=29

Figure 1.2

Patterns in which N_1=42 Without (a) and With (b)

Dependent Information in The Target Column

displays. Also a true optimum detector, which takes the inter-symbol dependencies into consideration, can distinguish between such displays extremely well.

1.3 The Present Investigation

This paper investigates the ability of human operators to make use of information presented by higher-order statistical processes, and the relation of the human to an optimum detector. As such, this effort represents an extension of the previously mentioned work to the more general case, and also answers some basic questions concerning the capabilities of the human to process dependent statistical information.

It should be pointed out that Julesz (5) has studied a different, but somewhat related, problem. Julesz was concerned with the ability of the human to discriminate between simultaneously presented visual fields of dots. The brightness of each point in his displays took on one of either 2,3, or 4 values and were determined by the output of a Markov chain. His investigations were concerned with finding specific visual properties of the display which allow human discrimination, in contrast to the present study which is concerned directly with the information content of the display and the ability of the human, as well as an optimum detector, to use different types of statistical information.

1.4 General Outline of Experiments

The experiments involved in this study may be grouped into three major classes. These experiments are discussed in general here to briefly outline the approach of the remainder of the paper, and will be presented in detail in the following chapters.

Experiment 1 was designed to answer three basic questions.

1. Are humans sensitive to information contained in the higher-

order (greater than first-order) statistics of a finite-valued, discrete information source?

2. If so, what is the approximate range of sensitivity, i.e., that region of stimulus intensity which does not lead to the two trivial detection probabilities of zero and one?
3. Within the range of sensitivity, does the human consistently favor one form of information over another?

Experiment 2 extends the domain of visual stimuli to a sub-set of the patterns generated by a stationary, first-order, binary Markov process. The questions asked in Experiment 2 are:

1. When using patterns which fall into overlapping classes (a pattern may exist in more than one class) does the human perform better or worse than with patterns from non-overlapping classes?
2. What is the form of the "operator noise" introduced in the visual detection process?

On the basis of the data obtained, the operator noise, as compared to a "noiseless" optimum detector, is determined, as well as the just noticable difference (j.n.d.) of the stimulus intensity.

Experiment 3 first provides a definition of amount of information, or "dissimilarity", contained in patterns in terms of independent and dependent components, and then goes on to discuss the question of the human's relative use of independence and dependent information when both are presented simultaneously. The questions specifically answered by Experiment 3 are:

1. What is the form of the change in operator's probability of correct detection when the relative amounts of independent

and dependent "dissimilarities" (information) in the displays are varied?

2. How does the performance of a human operator compare with that of a first-order detector and a true (Markov) optimum detector when varying amounts of component "dissimilarities" are presented?

Chapter 2

General Experimental Conditions and Apparatus

In the following chapters three experiments are discussed to answer the questions posed in Chapter 1. Throughout these experiments the same apparatus is used and certain psychophysical conditions remain constant. In this chapter these invariant properties of the experiments are discussed. Later, the specific details peculiar to each experiment are presented in greater detail.

The heart of the apparatus is a Digital Equipment Corp. PDP-5 digital computer - a flexible, small scale (4096 - 12 bit word core memory) general purpose machine. Other major elements of the system include a wide band (DC - 100kHz) Gaussian distributed noise generator, analog to digital converter, and Fairchild 737A 17 inch electrostatically deflected oscilloscope display (CRT). The computer in conjunction with the above equipment and miscellaneous external sweep and logic circuitry, is used to generate the displays under program control. In addition, the PDP-5 is used to control the sequencing of the experiments and to collect and process experimental data. Figure 2.1 shows a general block diagram of the system, and a more detailed description has been discussed in the literature (6,7).

The display consists of a 6 by 7 inch array of dots (64 by 84) on the face of the CRT. The points in all but one column of the display, the so-called "target column", are generated by a computer simulated statistically independent process with the probability of intensifying each point (corresponding to a binary "1") equal to the probability of not intensifying the point (a "0"). This is accomplished by independently sampling the noise generator at a slow 3kHz rate and converting the

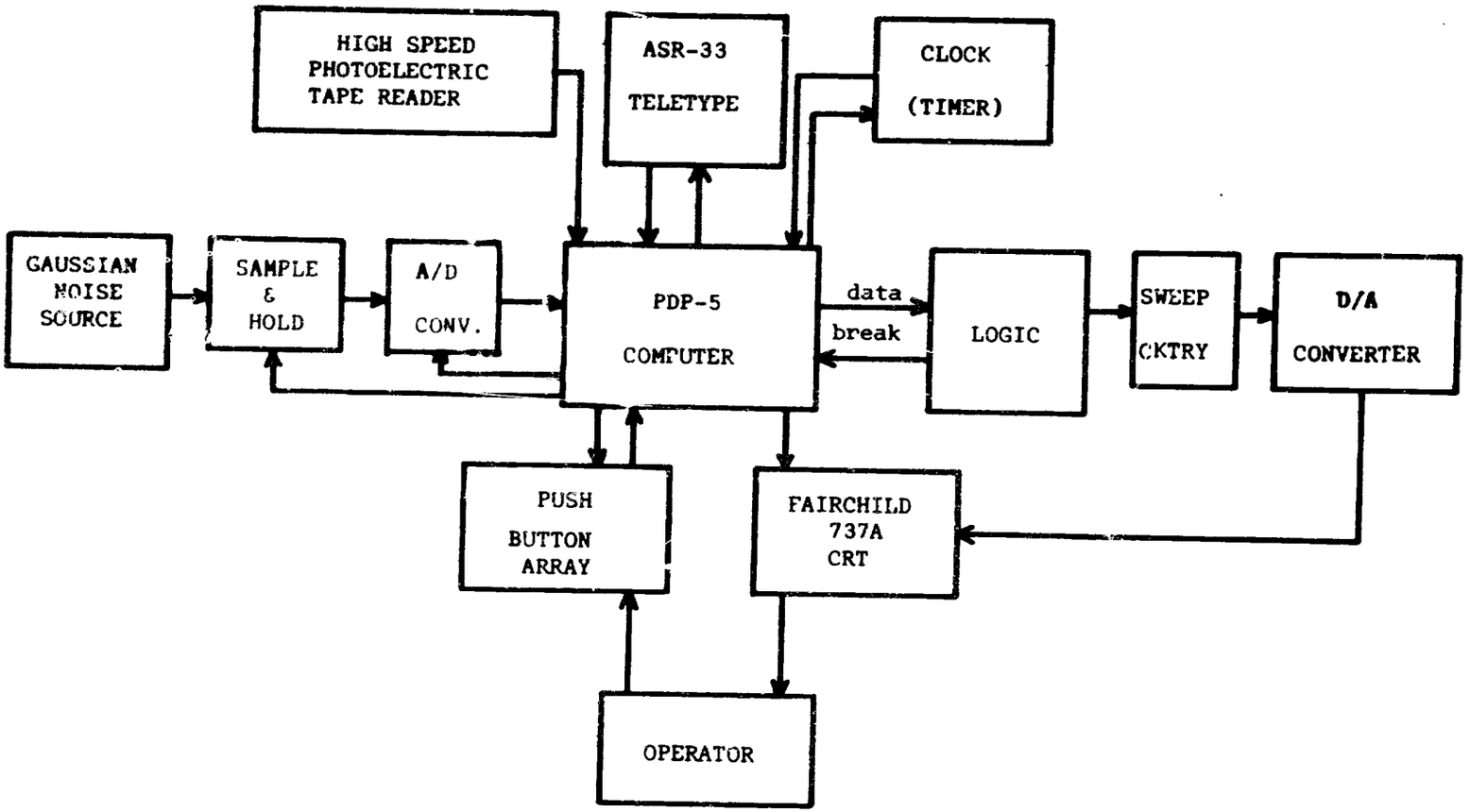


Figure 2.1

Block Diagram of Experimental Apparatus

resulting analog voltage into a 12 digit binary number, which is then "clipped" about its mean value generating a 0 or 1. Thus P(0)=P(1)= 1/2 with no inter-symbol dependencies in any columns except the target column. The statistical process used to generate the binary points in the target column depends on the particular experiment and is discussed in detail in the following chapters. Figure 2.2 shows a typical pattern as seen by the operator. The points are intensified at such a rate that no flicker is present, and markers are used above and below the target column to indicate its position to the operator.

The operator views the display through a hood which positions him 23 inches directly in front of the display. A small amount of light is shown around the edge of the display to eliminate any visual "burst" when the display comes on and goes off. The operator is allowed to control the brightness of the display to compensate for dark adaption. The operator's decisions are signalled to the computer by push buttons located in an array in front of him.

The display and operator are located in a 7 ft. high by 4 ft. wide by 6 ft. long darkened and soundproofed booth. The use of the previously mentioned hood, and the presence of nearly "white" background noise from a cooling fan isolate the operator from external stimuli and allow him to focus his full attention on the display screen.

In a typical session, the operator loads a program tape into the computer, adjusts the equipment, and enters the booth. Upon pressing a "start" button the first display appears. There is no time limit on how long he may view the display before making a decision, but he is asked to work as rapidly as he feels he can without diminishing confidence in his decisions.

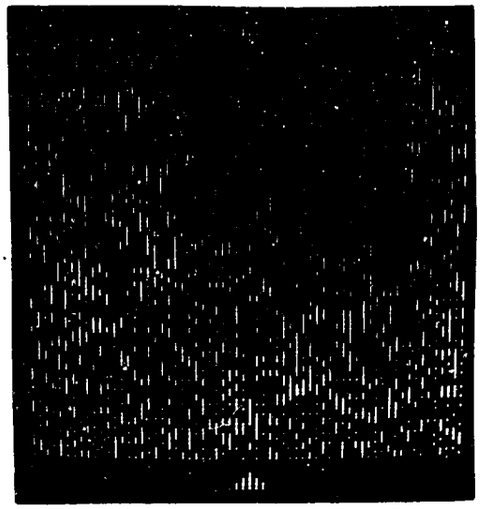
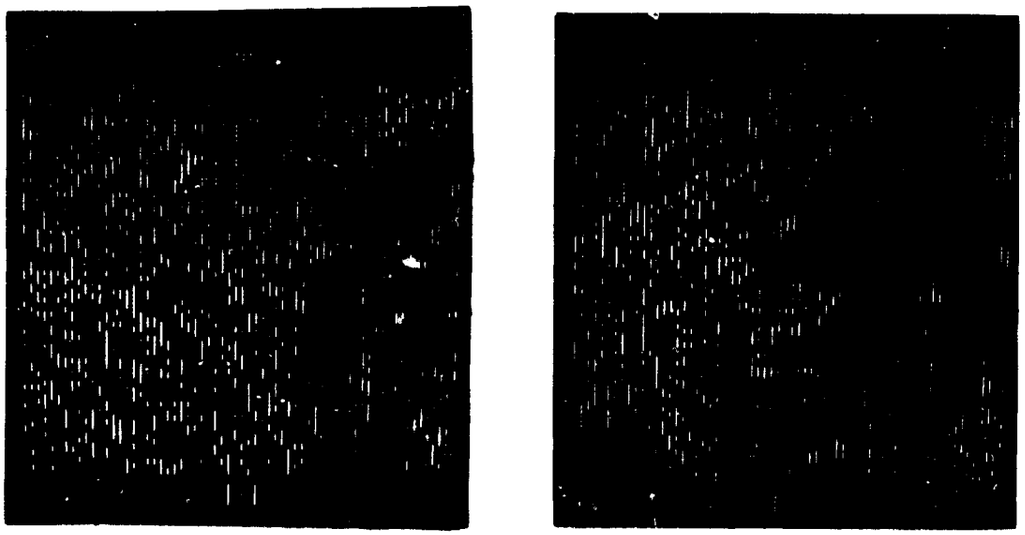


Figure 2.2

Typical Random Display



(a) (b)

Figure 2.3

Format of Feedback of Knowledge of Results

The operator's decision time is measured by a computer controlled clock, and recorded, along with his decision, when he presses a decision button. At this point in most of the experiments hit (H) or miss (M) information (and for the case of three choices of decision, the correct decision also) appears on the screen below (or above) the target column in place of the markers, as illustrated in Figure 2.3. This feedback of knowledge of results, is used as an immediate corrective factor to train the operator in the task which he is performing.

Between displays the screen is dark (except for the glow of the lights in the hood) for about two to four seconds (depending on the particular experiment) while the subsequent display is being generated. For any one experiment the display generation time is equalized for all types of displays which may be presented so that no clue as to the type of display can be obtained extraneously through this factor.

At the end of a session, consisting of either 100 or 150 trials, the display goes off and does not return. A tabulation of the data from the session is compiled by the computer, and is typed out on a teleprinter, as well as on paper tape for further processing.

Chapter 3

The Optimum Detector

3.1 Discussion of the Optimum Detector

Before discussing the experimental aspects of the thesis, it is helpful to develop a mathematical description of the statistically optimum detector. Knowledge of the form and capabilities of an "optimum" or "ideal" detector serves two purposes. First, the form of the optimum detector lends some insight into the possible factors affecting human detection capabilities. Second, the performance of an optimum detector provides a yard stick against which human performance may be compared.

Consider two information sources, S_1 and S_2, which generate discrete outputs at event times $\tau_1, \tau_2, \cdots \tau_i$. If one or the other of these sources is chosen at random, as depicted in Figure 3.1, and the output sequence $y_1 y_2 \cdots y_t$ observed, the problem which exists is to determine which source is the generating source. In the experiments which are discussed in the following chapters, this is the problem given to the subject.

Let H_1 and H_2 be the hypotheses that the output sequence $y_1 y_2 \cdots y_t$ was generated by S_1 and S_2 respectively. To simplify notation let $Y_{a,b}$ be the sequence of consecutive outputs $y_a y_{a+1} \cdots y_{b-1} y_b$ of length b-a+1, and let Y_a be the sequence of length one consisting of the single output symbol y_a.

An ideal detector (8) should calculate the likelihood ratio, $L(Y_{1,t})$, defined as

$$L(Y_{1,t}) = \frac{P(Y_{1,t}/H_1)}{P(Y_{1,t}/H_2)} \qquad 3.1.1$$

where $P(Y_{1,t}/H_i)$ is the probability of the output sequence of length t

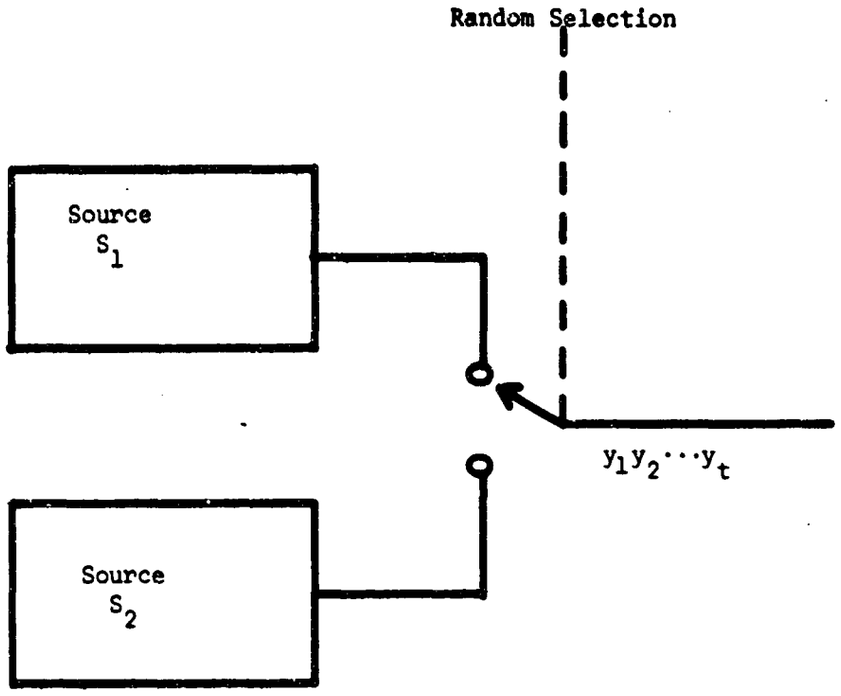


Figure 3.1

Random Selection of One of Two Sources

being generated, assuming hypothesis H_i is true. The likelihood ratio represents the confidence that S_1, rather than S_2, is the generating source. To make a decision, the likelihood ratio must be compared to a threshold, T, which is determined by the a priori probabilities $P(H_1)$ and $P(H_2)$ of H_1 and H_2, respectively, being true, and the relative costs of making each decision. The decision rule is:

$$\begin{aligned} L(Y_{1,t}) \geq T &: D_1 \quad (\text{source } S_1) \\ < T &: D_2 \quad (\text{source } S_2) \end{aligned} \qquad 3.1.2$$

where D_1 and D_2 are the respective decisions H_1 true and H_2 true. Let c_i (i=1,2) be the cost associated with making the incorrect decision D_i. Assume that no charge is made for correct decisions. When a priori probabilities $P(H_1)$ and $P(H_2)$ are known, the linear average cost function (Bayes Strategy) is:

$$\bar{c} = c_1 P(D_1/H_2)P(H_2) + c_2 P(D_2/H_1)P(H_1) \qquad 3.1.3$$

It has been shown (8) that the optimum decision threshold, which minimizes $\bar{c}$, is

$$T = \frac{P(H_2)c_1}{P(H_1)c_2} \qquad 3.1.4$$

3.2 Development of Optimum Detector for Markov Sources

In the previous section/the general form of a likelihood ratio decision strategy with a linear cost function was discussed. Here this technique is applied to the case in which S_1 and S_2 are Markov processes of order r_1 and r_2 respectively with identical output symbol sets $\{s_i; i=1,m\}$ consisting of m elements.

A basic property of an rth order Markov process is that the value

of the current output depends on only the past r outputs. Thus, the following conditional probability relation holds for all i:

$$P(y_i/y_1y_2\cdots y_{i-1}) = P(y_i/y_{i-r}y_{i-r+1}\cdots y_{i-1}) \qquad 3.2.1$$

Using the simplified notation introduced in the previous section, the above may be rewritten:

$$P(Y_i/Y_{1,i-1}) = P(Y_i/Y_{i-r,i-1}) \qquad 3.2.2$$

Recalling the form of the optimum detector expressed in relation 3.1.2, and making use of the above relation, the optimum detector for a string of symbols emmited by one of two Markov sources may be written:

$$\frac{P(Y_{1,r}/H_1)\,P(Y_{r+1,t}/Y_{1,r};H_1)}{P(Y_{1,r}/H_2)\,P(Y_{r+1,t}/Y_{1,r};H_1)} \gtrless \frac{P(H_2)c_1}{P(H_1)c_2} \quad \begin{matrix} : D_1 \\ < \quad " \quad : D_2 \end{matrix} \qquad 3.2.3$$

This expression may be further expanded into the form:

$$\frac{P(Y_{1,r_1}/H_1)\,P(Y_{r_1+1}/Y_{1,r_1};H_1)\,P(Y_{r_1+2}/Y_{2,r_1+1};H_1)\cdots P(Y_t/Y_{t-r_1,t-1};H_1)}{P(Y_{1,r_2}/H_2)\,P(Y_{r_2+1}/Y_{1,r_2};H_2)\,P(Y_{r_2+2}/Y_{2,r_2+1};H_2)\cdots P(Y_t/Y_{t-r_2,t-1};H_2)}$$

$$\gtrless \frac{P(H_2)c_1}{P(H_1)c_2} \quad : D_1 \qquad 3.2.4$$

Otherwise : D_2

Each conditional sequence probability in expression 3.2.4 represents the probability that a particular output, y_i, will take on some particular value, given that the past r_k outputs have taken on particular values and that one of the two hypotheses is true. Since there are m values which each sequence of length one may take on, and m^{r_k} possible sequences of length r_k, there are $m \cdot m^{r_k} = m^{r_k+1}$ possible values for each of the $t-r_k$ conditional probabilities which must be considered in both the

numerator and denominator of expression 3.2.4. Let these conditional probabilities (and their associated r_k+1 length sequences) be ordered as follows:

$$
\begin{array}{ll}
P_{k,1}=P(S_1/S_1\cdots S_1,H_k) & Y^{k,1}=S_1S_1\cdots S_1S_1 \\
P_{k,2}=P(S_2/S_1\cdots S_1,H_k) & Y^{k,2}=S_1S_1\cdots S_1S_2 \\
\cdot & \cdot \\
\cdot & \cdot \\
\cdot & \cdot \\
P_{k,m}=P(S_m/S_1\cdots S_1,H_k) & Y^{k,m}=S_1S_1\cdots S_1S_m \\
P_{k,m+1}=P(S_1/S_1S_1\cdots S_2,H_k) & Y^{k,m+1}=S_1S_1\cdots S_2S_1 \\
\cdot & \cdot \\
\cdot & \cdot \\
\cdot & \cdot \\
P_{k,2m}=P(S_m/S_1S_1\cdots S_2,H_k) & Y^{k,2m}=S_1S_1\cdots S_2S_m \\
\cdot & \cdot \\
\cdot & \cdot \\
\cdot & \cdot \\
P_{k,m^{r_k+1}}=P(S_m/S_m\cdots S_m,H_k) & Y^{k,m^{r_{k+1}}}=S_mS_m\cdots S_m
\end{array}
\qquad 3.2.5
$$

for k=1,2

This represents a natural ordering of the r_k+1 length sequences $Y^{k,i}$ with the last (right most) symbol running through its m possible values before the left adjacent symbol is incremented. The probabilities $P_{k,i}$ are just those corresponding to the conditional sequences associated with the $Y^{k,i}$. Note the following relations between these probabilities:

$$\sum_{j=0}^{m} P_{k,(m^{r_k}.j)+w} = P(S_w/H_k) \qquad 3.2.6a$$

$$\sum_{j=1}^{m^{r_k+1}} P_{k,j} = 1 \qquad 3.2.6b$$

for all $1 \leq w \leq m$; $k = 1,2$.

As an example of the above ordering consider two Markov sources, S_1 and S_2, with binary output sets. Let S_1 be a first-order process (r_1=1) and S_2 be of second-order (r_2=2).

The following conditional probabilities and sequences must be defined.

$P_{1,1} = P(0/0,H_1)$ $\quad Y^{1,1}=00$

$P_{1,2}=P(1/0,H_1)$ $\quad Y^{1,2}=01$

$P_{1,3}=P(0/1,H_1)$ $\quad Y^{1,3}=10$

$P_{1,4}=P(1/1,H_i)$ $\quad Y^{1,4}=11$

$P_{2,1}=P(0/00,H_2)$ $\quad Y^{2,1}=000$

$P_{2,2}=P(1/00,H_2)$ $\quad Y^{2,2}=001$

$P_{2,3}=P(0/01,H_2)$ $\quad Y^{2,3}=010$

$P_{2,4}=P(1/01,H_2)$ $\quad Y^{2,4}=011$

$P_{2,5}=P(0/10,H_2)$ $\quad Y^{2,5}=100$

$P_{2,6}=P(1/10,H_2)$ $\quad Y^{2,6}=101$

$P_{2,7}=P(0/11,H_2$ $\quad Y^{2,7}=110$

$P_{2,8}=P(1/11,H_2)$ $\quad Y^{2,8}=111$

Let $n_{k,j}$ be the number of times the j^{th} ordered sequence, $Y^{k,j}$, appears in the output sequence $Y_{1,t}$; then $n_{k,j}$ conditional probabilities of expression 3.2.4 take on the value $P_{k,j}$, and expression 3.2.4 becomes:

$$\frac{P(Y_{1,r_1}/H_1)\prod_{j=1}^{m} P_{1,j}^{n_{1,j}}}{P(Y_{1,r_2}/H_2)\prod_{j=1}^{m} P_{2,j}^{n_{2,j}}} \gtrless \frac{P(H_2)c_2}{P(H_1)c_2} \quad : D_1 \qquad 3.2.7$$

otherwise : D_2

Note that there will be at most t different sequences (i.e., of length one), so at most t different $n_{k,j}$ are different from zero. From this point on the threshold, T, will be taken as unity since this is the only case which will be discussed in later chapters. Normally the logarithm of this expression is taken, in which case we have:

$$\log P(Y_{1,r_1}/H_1) + \sum_{j=1}^{m} n_{1,j} \log P_{1,j} \gtrless \log P(Y_{1,r_2}/H_2) + \sum_{j=1}^{m} n_{2,j} \log P_{2,j} \quad : D_1 \qquad 3.2.8$$

otherwise : D_2

which is the final form for the general optimum detector when a sequence of output symbols may have been generated by one of two very general Markov processes.

In evaluating expressions 3.2.7 and 3.2.8 one should notice that it is necessary to count the number of occurrences of each of m^{r_k+1} sequences of length r_k+1 as well as making note of the exact form of the initial sub-sequence Y_{1,r_k}. The task of the optimum detector may be very greatly reduced(especially for large m and/or r_k) if some approximations are made. These approximations are introduced and a more useful form of the optimum detector is developed in the next section.

3.3 Simplification of the Form of Optimum Detector

Expression 3.2.8 may be greatly simplified if the following approximations are made. First, if r_k, the order of the Markov process, is much less than the length of the observed sequence, $Y_{1,t}$, the probability

of the initial sub-sequence Y_{1,r_k}, has little effect on the overall sequence probability. The log of the probability of the initial sub-sequence is small and of the same order of magnitude under the assumption of each hypothesis; therefore, the first term on each side of expression 3.2.8 may be dropped with little loss of accuracy. Expression 3.2.8 becomes:

$$\sum_{j=1}^{m^{r_1+1}} n_{k,j} \log P_{1,j} \gtrless \sum_{j=1}^{m^{r_2+1}} n_{k,j} \log P_{2,j} \quad : D_1 \qquad 3.3.1$$

$$\text{otherwise} \quad : D_2$$

The complexity which remains in relation 3.3.1 is due to the fact that evaluation of the expression requires observation of the frequency counts, $n_{k,j}$, of all m^{r_k+1} sequences of length r_k+1. These sequences are not independent however, and it is possible to represent the frequency count of many of these sequences as a linear combination of some smaller "basis" set of frequency counts. The approach used here is similar to that presented in Booth (9), for determining a minimal generator set of a random process. However, some modifications are necessary since we are dealing here with actual frequency counts and not the underlying probability structure.

Consider a Markov process of first-order (r=1) with two possible output symbols ($\{s_i; i=1,2\}=\{s_1,s_2\}=\{0,1\}$). There are four sequences of length r+1=2; these are:

$$Y^1 = 00$$
$$Y^2 = 01$$
$$Y^3 = 10 \qquad 3.3.2$$
$$Y^4 = 11$$

However, certain constraints exist on the number of these sub-sequences which may exist in a longer sequence, $Y_{1,t}$, of length t. If the symbols of $Y_{1,t}$ are considered in groups of two, they may be listed as:

$$\begin{array}{l} y_1y_2 \\ y_2y_3 \\ y_3y_4 \\ \cdot \\ \cdot \\ \cdot \\ y_{t-1}y_t \end{array} \qquad 3.3.3$$

and each of these pairs is one of the sub-sequences listed in 3.3.2. This makes up a set, $\{Y_{i,i+1};i=1, t-1\}$ each element of which is one of the sub-sequences of expression 3.3.2. Observe that the first element of each of the sequences listed in expression 3.3.3 when strung together form the sequence $Y_{1,t-1}$. Thus, the number of sequences of the set $\{Y_{i,i+1};i=1,t-1\}$ which begin with a 1 (i.e., $y_i = s_2 = 1$) make up an approximation of the number of 1's in $Y_{1,t}$.

Denote the number of sequences of $\{Y_{i,i+1};i=1,t-1\}$ which take on values Y^1, Y^2, Y^3, and Y^4 (of expression 3.3.2) by $N_{00}(Y_{1,t}), N_{01}(Y_{1,t}), N_{10}(Y_{1,t})$, and $N_{11}(Y_{1,t})$ respectively. Let $N_0(Y_{1,t})$ and $N_1(Y_{1,t})$ be the number of symbols (i.e., sequences of length one) of $Y_{1,t}$ which take on values s_1 (i.e., 0) and s_2(i.e., 1) respectively. Further, let $N_0(Y_k)$ and $N_1(Y_k)$ be a 1 if and only if Y_k is a 0 and a 1 respectively, and let N be the number of symbols in the sequence $Y_{1,t}$. The following constraints exist:

$$\begin{aligned} N_{00}(Y_{1,t}) + N_{01}(Y_{1,t}) &= N_0(Y_{1,t}) - N_0(Y_t) \\ N_{10}(Y_{1,t}) + N_{11}(Y_{1,t}) &= N_1(Y_{1,t}) - N_1(Y_t) \\ N_{00}(Y_{1,t}) + N_{10}(Y_{1,t}) &= N_0(Y_{1,t}) - N_0(Y_1) \\ N_{01}(Y_{1,t}) + N_{11}(Y_{1,t}) &= N_1(Y_{1,t}) - N_1(Y_1) \end{aligned} \qquad 3.3.4$$

The second term on the right side of each of the expressions 3.3.4 is either one or zero. Thus, it may be dropped completely in most cases with little loss of accuracy. Note, also, that $N=N_0(Y_{1,t}) + N_1(Y_{1,t}) = N_{00}(Y_{1,t})+N_{01}(Y_{1,t})+N_{10}(Y_{1,t})+N_{11}(Y_{1,t})$.

From expression 3.3.4 and the immediately preceeding relation we may write the following relations:

$$N_1(Y_{1,t}) = N_1(Y_{1,t})$$
$$N_0(Y_{1,t}) = N - N_1(Y_{1,t})$$

3.3.5

$$N_{11}(Y_{1,t}) = N_{11}(Y_{1,t})$$
$$N_{10}(Y_{1,t}) = N_1(Y_{1,t}) - N_{11}(Y_{1,t})$$
$$N_{01}(Y_{1,t}) = N_1(Y_{1,t}) - N_{11}(Y_{1,t})$$
$$N_{00}(Y_{1,t}) = N - 2N_1(Y_{1,t}) + N_{11}(Y_{1,t})$$

Note that each of the four frequency counts of sequences of length two in the above relations has been written as a linear combination of the "basis" counts $\{N, N_1, N_{11}\}$. This is not the only "basis" which could have been chosen; among the others are:

$$\{N, N_0, N_{00}\}$$
$$\{N, N_1, N_{01}\} \qquad 3.3.6$$
$$\{N, N_0, N_{11}\}$$

In this case there are eight different basis sets which may be chosen.

Extending the above reasoning to the general case of an r^{th} order Markov process with m symbols, one may choose a basis set of frequency counts by observing the constraints on equations 3.3.4. For sequences of length one we have the constraints:

$$N_{s_1}(Y_{1,t}) + N_{s_2}(Y_{1,t}) + \cdots + N_{s_m}(Y_{1,t}) = N(Y_{1,t}) \qquad 3.3.7$$

Thus, if m-1 of the frequency counts of sequences of length one are known (and N is known) equation 3.3.7 says that the m^{th} frequency count may be uniquely determined. For sequences of length two the following constraints hold:

$$\sum_{i=1}^{r} N_{s_i s_j}(Y_{1,t}) = N_{s_j}(Y_{2,t}) \simeq N_{s_j}(Y_{1,t}) \qquad 3.3.8$$

and $\sum_{i=1}^{m} N_{s_j s_i}(Y_{1,t}) = N_{s_j}(Y_{1,t-1}) \simeq N_{s_j}(Y_{1,t})$ for j = (1,m)

The approximation holds only if $N >> 1$. These 2m equations involve m^2 unknown frequency counts for sequences of length two. One equation is a linear combination of the other 2m-1 because of restriction 3.3.7. There are $m^2-(2m-1) = (m-1)^2$ frequency counts for sequences of length two which must be selected according to expression 3.3.8.

Considering frequency counts of longer and longer sequences, up to length r+1, we see that there will be $2m^r$ constraints on the m^{r+1} frequency counts of sequences of length r+1 of the form:

$$\sum_{i=1}^{m} N_{s_i s_{j_1} s_{j_2} \cdots s_{j_r}}(Y_{1,t}) = N_{s_{j_1} s_{j_2} \cdots s_{j_r}}(Y_{2,t}) \simeq N_{s_{j_1} \cdots s_{j_r}}(Y_{1,t}) \qquad 3.3.9$$

and

$$\sum_{i=1}^{m} N_{s_{j_1} s_{j_2} \cdots s_{j_r} s_i}(Y_{1,t}) = N_{s_{j_1} s_{j_2} \cdots s_{j_r}}(Y_{1,t-1}) \simeq N_{s_{j_1} \cdots s_{j_r}}(Y_{1,t})$$

But there will be m^{r-1} restrictions on frequency counts of sequences of length r. There will be, then, $m^{r+1}-(2m^r-m^{r-1}) = m^{r-1}(m-1)^2$ frequency counts of sequences of length r+1 which can be selected independently. There are a total of $(m-1)+(m-1)^2+\cdots+(m-1)^2 m^{r-1} = (m-1)\, m^r$ basis frequency counts necessary to approximate all the frequency counts of sequences

of length r+1. In addition, the number of symbols, N, must be known.

Let us return now to the notation adopted in section 3.2; in particular let $n_{k,j}$ be the number of times the k,j^{th} ordered sequence (see 3.2.5) appears in $Y_{1,t}$. Call the elements of the set of basis frequency counts $F_{k,i}$. There will be at most $\{m^{r_k}(m-1) + 1\}$ $F_{k,i}$'s needed to specify frequency counts of sequences of length r_k+1. Any one of these frequency counts, $n_{k,j}$ may be expressed as a linear combination of the $m^{r_k}(m-1) + 1$ basis counts.

$$n_{k,j} = \sum_{i=0}^{m^{r_k}(m-1)} f_{k,j,i} F_{k,i} \qquad 3.3.10$$

for all k=1,2; j=(1,m)

where $F_{k,0}$ is defined to be N, and $f_{k,j,i}$ is the integer weighting factor associated with the k,j^{th} ordered sequence and the i^{th} basis frequency count.

The optimum detector (3.2.8) now becomes: 3.3.11

$$\sum_{j=1}^{m^{r_1+1}} \left[\sum_{i=0}^{m^{r_1}(m-1)} f_{1,j,i} F_{1,i} \right] \log P_{1,j} \gtrless \sum_{j=1}^{m^{r_2+1}} \left[\sum_{i=0}^{m^{r_2}(m-1)} f_{2,j,i} F_{2,i} \right] \log P_{2,j}$$

$$: D_1$$

otherwise : D_2

Interchanging the order of summation, and regrouping gives: 3.3.12

$$\sum_{i=0}^{m^{r_1}(m-1)} F_{1,i} \log \prod_{j=1}^{m^{r_1+1}} (P_{1,j})^{f_{1,j,i}} \gtrless \sum_{i=0}^{m^{r_2}(m-1)} F_{2,i} \log \prod_{j=1}^{m^{r_2+1}} (P_{2,j})^{f_{2,j,i}} \quad : D_1$$

otherwise : D_2

Note that the form 3.3.12 involves observation of only $m^{r_k}(m-1)$ frequency counts of the sequences $Y_{1,t}$ to evaluate the summation on each side. This allows a saving of $m^{r_k+1} - m^{r_k}(m-1) = m^{r_k}$ frequency counts over the use of form 3.3.1. For large m and/or r_k this saving can be substantial.

3.4 Special Forms of the Optimum Detector

In this section a few special forms of the optimum detector 3.3.12 are developed. These forms will be used in later chapters when the Markov optimum detector is compared to the human operator.

First, for the case of $r_1=r_2=r$, the optimum detector reduces to:

$$\sum_{i=0}^{m^r(m-1)} F_i \log \prod_{j=1}^{m^{r+1}} \left(\frac{P_{i,j}}{P_{2,j}}\right)^{f_{j,i}} \geq 0 \quad : \quad D_1$$

otherwise: D_2 3.4.1

If one of the sources, S_2 say, is actually a statistically independent process (i.e., r=0), $P(Y_i/Y_{1,i-1})=P(Y_i)$ for all i=1,2,···, and expression 3.3.12 reduces to:

$$\sum_{i=0}^{m^{r_1}(m-1)} F_{1,i} \log \prod_{j=1}^{m^{r_1+1}} (P_{1,j})^{f_{1,j,i}} \geq \sum_{u=1}^{m} N_{s_u}(Y_{1,t}) \log P(s_u) \quad : \quad D_1$$

otherwise : D_2 3.4.2

where $N_{s_u}(Y_{1,t})$ is the number of times the symbol s_u appeared in the observed sequence $Y_{1,t}$, and $P(s_u)$ in the probability that y_i takes on value s_u (for all i=(1,t)). Note that the left side of 3.4.2 was derived from the approximation that t >> r and the probability of the initial subsequence $Y_{1,r}$ could be dropped. This approximation has more effect in 3.4.2 since no quantity of similar magnitude is being dropped on the right side. A better approximation would be obtained

if the first r symbols were not considered in evaluating the right side, i. e., let $Y_{1,t} \rightarrow Y_{r+1,t}$ in 3.4.2.

Specifically, for S_1 a first-order Markov process with a binary symbol set $\{S_j;j=1,2\}=\{S_1,S_2\}=\{0,1\}$ and S_2 a binary statistically independent process with $P(0) = P(1) = 1/2$, the optimum detector may be expressed in approximate form as:

$$\sum_{i=0}^{2} F_{1,i} \log \prod_{j=1}^{4} (P_{1,j})^{f_{1,j,i}} \gtrless (t-1) \log \left(\frac{1}{2}\right) : D_1 \text{ (Independent)} \qquad 3.4.3$$

$$\text{otherwise} : D_2 \text{ (Markov)}$$

where the $F_{1,i}$ (and $f_{1,j,i}$) are chosen as in section 3.3. For the "basis" set mentioned in section 3.3, one specific form of expression 3.4.3 is:

$$N \log P(0/0) + N_1 \log \frac{P(1/0)P(0/1)}{P^2(0/0)} + N_{11} \log \frac{P(0/0)P(1/1)}{P(1/0)P(0/1)} \gtrless (t-1)\log \frac{1}{2} \qquad 3.4.4$$

$$: D_1$$

$$\text{Otherwise} : D_2$$

where $\{F_{1,i}\}= \{N,N_1N_{11}\}$

and

$$f_{1,j,i} = \begin{bmatrix} +1 & -2 & +1 \\ 0 & +1 & -1 \\ 0 & +1 & -1 \\ 0 & 0 & +1 \end{bmatrix}$$

$P_{2,1} = P(0/0)$
$P_{2,2} = P(1/0)$
$P_{2,3} = P(0/1)$
$P_{2,4} = P(1/1)$

3.5 Implications of the Optimum Detector

In the preceeding sections it has been shown that it is possible to formulate the design of an optimum detector which makes use of higher-order information. Specifically, for a first-order binary Markov process it was shown that the optimum detector results in a weighted

summation of frequency counts of sequences of length one and two. It is reasonable to ask if the human operator can also extract this information, and, if so, to what extent. Also, does the human use dependent information in a manner similar to the optimum detector, or does he use different cues.

After determining, in the next chapter, the range of human sensitivity to dependent information, experiments are discussed which answer the above question.

Chapter 4

Experiment 1 - Basic Questions

4.1 Introduction to Experiment 1

Before considering some of the detailed aspects of the effects of inter-symbol dependencies on human visual detection capabilities, it is necessary to determine the range of dependencies to which the human is sensitive, and whether or not he favors certain types of dependencies over others. Julesz's work in visual discrimination (5) has shown that humans more easily discriminate between two visual fields when the border exhibits a "connectivity " property. In other words, if the human can subjectively "connect" a "line" of equal brightness levels, his discrimination is facilitated. It was thought that perhaps the subjects in the present investigation might, on this basis, favor one type of dependency over another.

Specifically, Experiment 1 was designed to answer three fundamental questions which provide some basic insight into human performance in this particular area. It also provides the information necessary for the design of later experiments.

1. Is a human inherently sensitive to information provided by the dependencies between consecutive symbols of a visual display? In other words, without previous training can a subject learn to correctly identify displays which differ only in their inter-symbol dependencies when no knowledge of results is provided to reinforce or modify the subject's performance.
2. When feedback of knowledge of results is provided does the human learn to detect information provided by inter-symbol

dependencies, and does his performance improve to some steady-state? If so, what is the level of this steady-state performance?

3. What range of dependencies leads to non-trivial (other than zero and one) detection probabilities? What is the range of human sensitivity where more detailed investigations should be concentrated?

4.2 Design of Experiment 1

To answer the above three questions, the following experiment was performed. Using the general display scheme outlined in Chapter 2, displays were presented to subjects for classification into one of three groups. On each trial the subject had equal chances (1/3) of viewing any one of three types of patterns. Each pattern contained 63 columns of background noise consisting of 84 points in each column which were generated by a simulated, statistically independent, process with $P(0) = P(1) = 1/2$. The statistics of these 63 noise background columns remained constant over all trials. The target column, located near the center of the display, also contained 84 points, but was chosen to possess very specific properties. On every trial the number of 1's and 0's in the target column was each exactly 42. This is one half of the total number of points,and also represents the expected value of the number of 1's and 0's in the noise background columns. The number of 11's(and 00's) is the cue on which the subject based his decision, and was set randomly at one of three levels, $N/4 = 21$, $21+\delta$, and $21-\delta$. The parameter δ was fixed for each session of 150 trials, and took on values of either 2,4, or 6 depending on the particular experiment. Since

the number of 11's and 00's was increased or decreased by an amount δ, the number of 01's and 10's had to be decreased or increased respectively to maintain the same total number of points in each column. The properties of these target columns are admittedly very special and are not related specifically to any statistical process, but are, rather, of a deterministic nature. These types of target columns were used, however, because they were sufficient to answer the questions at hand, and were simple to generate. Once they were generated and stored on paper tape they were available for all experiments with different subjects. Figure 4.1 shows some typical displays with $N_{11}=21 \pm \delta$, for δ of 2, 4, and 6.

The patterns used in this experiment are deterministic in the sense that an optimum detector may employ a decision rule which leads to a detection probability of 1. As demonstrated in Figure 4.2, the probability density function of the number of 11 sequences in the three types of displays is simply three delta functions with magnitudes of 1/3 each. Placement of decision thresholds T_1 and T_2 between the peaks of the density function leads to an optimum detector with perfect performance. Since N_{11} must vary by at least one count (i.e., δ is an integer: $\delta \geq 1$) placement of decision thresholds at $T_1=N/4 - 1/2$ and $T_2=N/4 + 1/2$ leads to perfect detection for any δ. The problem under investigation is the determination of the range of δ to which the human is sensitive and whether or not he consistently favors an increase or a decrease in N_{11} over the opposite situation.

4.3 Results of Experiment 1

Since there were no data available on human performance in a visual detection task with statistically generated dependent symbols,

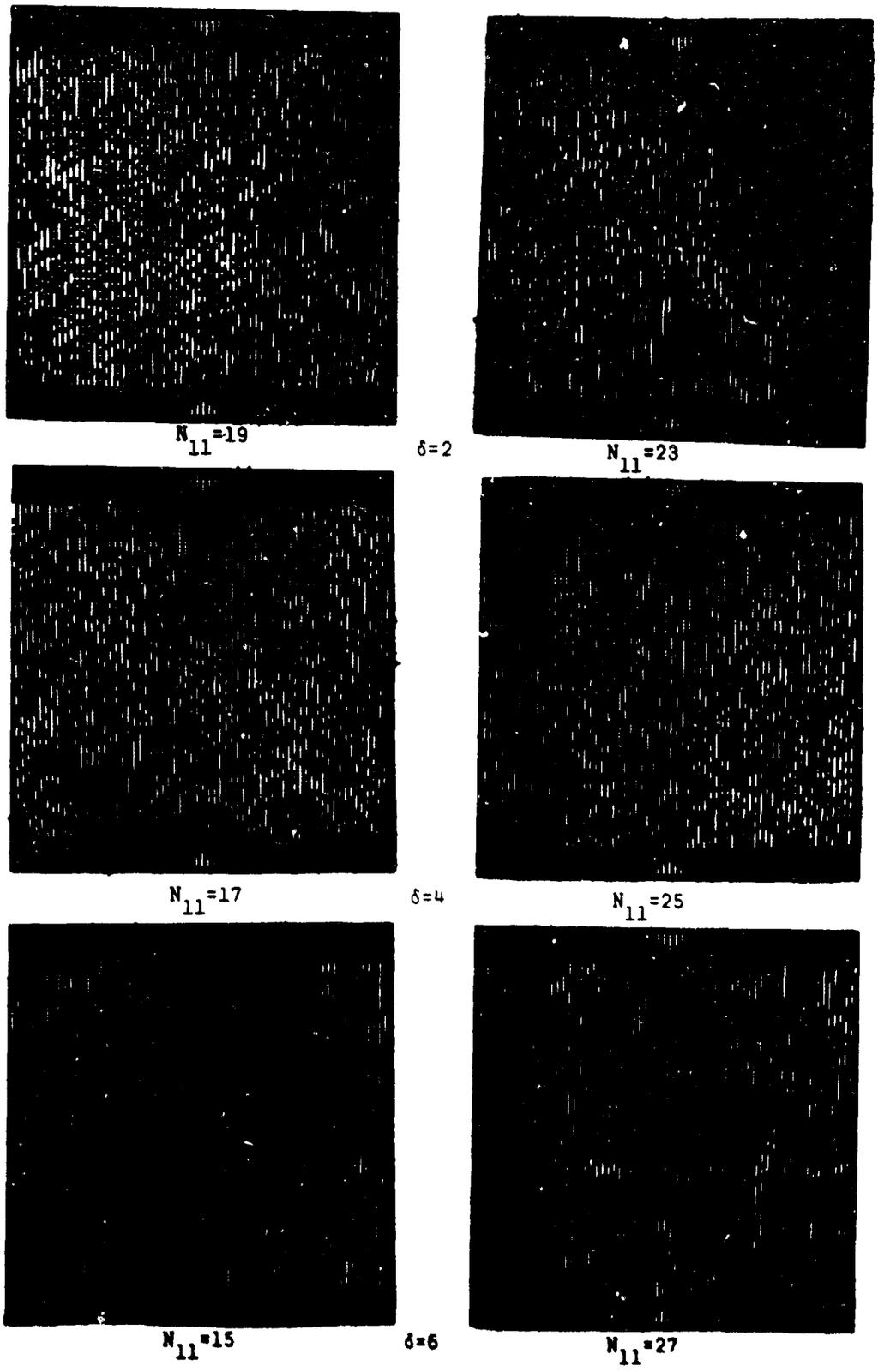


Figure 4.1

Typical Displays, Experiment 1, Conditions: δ= 2,4,6.

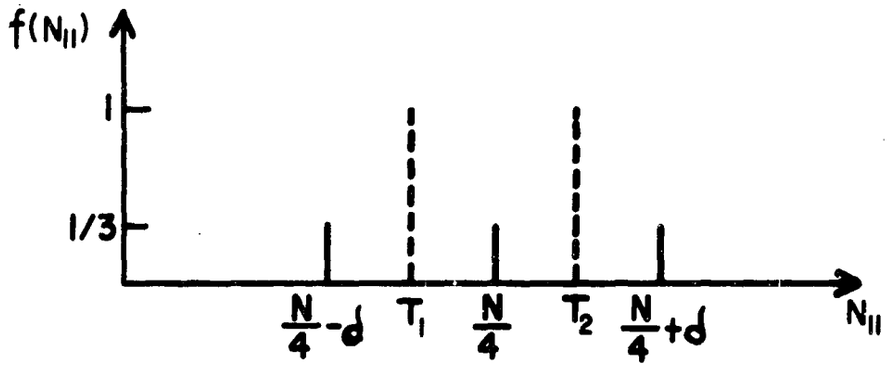


Figure 4.2

Probability Density Function of N_{11} for Experiment 1

$$f(N_{11})=(1/3)[\delta(N_{11}-(N/4-\delta))+\delta(N_{11}-N/4)+\delta(N_{11}-(N/4+\delta))]$$

Optimum Decision Rule:

$N_{11}<T_1$: D_1 (choose hypothesis H_1: $N_{11}=\frac{N}{4}-\delta$)

$T_1 \leq N_{11} \leq T_2$: D_2 (choose hypothesis H_2: $N_{11}=\frac{N}{4}$)

$T_2 < N_{11}$: D_3 (choose hypothesis H_3: $N_{11}=\frac{N}{4}+\delta$)

Detection Probabilities:

$$P(D_1/H_1) = P(N_{11}<T_1/H_1)= \int_{-\infty}^{T_1} f(N_{11}/H_1)dN_{11}=\int_{-\infty}^{T_1} \delta(N_{11}-(\frac{N}{4}-\delta))dN_{11}=1$$

if $T_1 > \frac{N}{4}-\delta$

$$P(D_2/H_2)=P(T_1 \leq N_{11} \leq T_2/H_2) = \int_{T_1}^{T_2} f(N_{11}/H_2)dN_{11}=\int_{T_1}^{T_2}\delta(N_{11}-\frac{N}{4})dN_{11}=1$$

if $T_1 \leq \frac{N}{4} \leq T_2$

$$P(D_3/H_3)=P(T_2 < N_{11}/H_3) = \int_{T_2}^{\infty} f(N_{11}/H_3)dN_{11}=\int_{T_2}^{\infty} \delta(N_{11}-(N/4+\delta))dN_{11}=1$$

if $T_2 < \frac{N}{4}+\delta$

the first problem was to determine the range of human sensitivity to dependent information so that further experiments could be meaningfully designed.

In the first phase of Experiment 1 two subjects were run under various conditions on δ without any previous discussion of the type of patterns which might appear and without any feedback of knowledge of results. Both subjects for this phase had no previous display experience.

First, Subject A was presented for one session with displays in which $\delta = 2$. He was told that the patterns would fall into three classes, and was instructed to try to classify the patterns consistently by pressing one of three buttons after each display appeared. He was told to take his time and to look over the display carefully. Subject A was also told that the differences between the three types of patterns would occur in the target column, which was marked above and below by pointers. He was not given any indication of the way in which the pattern classes differed.

After 150 trials (50 of each type of display) Subject A showed no consistent decision strategy related to the number of 11 sequences in the target column. His overall stimulus-response matrix was:

		STIMULUS		
		1	2	3
RESPONSE	"1"	.133	.133	.107
	"2"	.113	.127	.113
	"3"	.093	.073	.113

$\delta = 2$, no training, no feedback

It should be pointed out that this form of stimulus-response (S-R) matrix contains elements which represent the relative frequency of the joint occurrence "response-i to stimulus-j". The sum of all elements is one; each column sum is the relative frequency of the occurrence of that particular stimulus; and the row sums indicate the portion of the subject's responses which were of that certain type. The sum of the diagonal elements represents the relative frequency of correct decisions. In this case the probability of correct classification was .33 which does not differ significantly from a chance value of 1/3.

Since performance was so poor at $\delta = 2$, the next level of stimulus investigated was $\delta = 6$. Under this condition the same subject immediately began to classify the three types of patterns consistently. His overall correct detection probability rose to about .68, the actual S-R matrix being:

STIMULUS

RESPONSE	1	2	3
"1"	.32	.153	0
"2"	.006	.06	.24
"3"	.006	.12	.093

$\delta = 6$ no feedback

The subject had obviously chosen to call stimulus-3 by the name "type-2". Thus interchanging rows 2 and 3 "corrects" the subject's naming procedure to that of the experimenter.

		STIMULUS		
		1	2	3
RESPONSE	1	.32	.153	0
	2	.006	.12	.093
	3	.006	.06	.24

"corrected" S-R matrix

Detection Probability = 0.680

The subject's performance on stimulus-2 was rather poor; he had trouble deciding whether to make response-1 or response-2. The S-R matrix does, however, clearly reflect an ability to extract information provided by a difference in the number of second order sequences only. Recall that $N_1=N_0=42$ for all target columns. The answer to the first question posed in section 4.1 is that a human is inherently sensitive to higher order information in this task provided that the information is sufficient to separate displays by at least five to six counts of sequences of length two.

Before commencing with sessions in which knowledge of results was provided after each trial, a second naive subject was run under conditions similar to the above. However, prior to running, subject B was informed of the display generation procedure and the characteristics of the various patterns which would appear. It was explained to him that $N_1=N_0=42$ in the target column and that the background was random with an expected value of the number of 1's (and 0's) of 42, but that the target column would have either 21, 15, or 27 11 and (00) sequences,

while the average number of these sequences in the background would be 21. No knowledge of results was provided during the 150 trials. The subject's resulting S-R matrix was:

		STIMULUS		
		1	2	3
RESPONSE	1	.263	.04	0
	2	.07	.26	.053
	3	0	.033	.28

δ=6 Initial Training, no feedback

Setection Probability = .803

Subject B performed with a probability of correct decision of about 8/10. Clearly detection of patterns with δ=6 is a relatively simple task once the subject learns what to look for. The question which arises is to what level will a subject's performance rise when he is given extended practice and knowledge of results? What are the subjects' "steady-state" capabilities after learning dynamics have died out? Phase two of Experiment 1 provides an answer to this second question.

Phase two of Experiment 1 was identical to phase one except that the generation procedure and properties of the patterns were described in detail to all subjects prior to the first session. Knowledge of results was provided after each decision by changing the pointer below the target column into an "H" for "hit" or "M" for "miss", and the upper pointer into the correct pattern type, "1", "2", or "3". Three paid subjects, in addition to the author, participated in this experiment.

Subjects A and B, male undergraduate engineering students, were also subjects A and B in phase 1. Subject C, a female graduate student, had had no previous display experience. The author may be considered to be subject D.

Under condition $\delta = 2$ subject detection probability averaged over all classes of patterns and the three subjects (B, C, and D) participating was 0.576, well above a chance level. The overall S-R matrix based on pooled data from 2550 trials of three subjects' later runs reflect an ability to learn to detect patterns differing only by two second order counts.

		STIMULUS		
		1	2	3
RESPONSE	1	.197	.067	.017
	2	.103	.169	.106
	3	.033	.097	.210

Pooled Data, 2550 Trials, $\delta = 2$

Detection Probability = .576

In this experiment all subjects favored stimulus-3; that is they had a bias toward making response-3. This effect diminished somewhat in later sessions but never disappeared completely. If this increased ability to detect stimulus-3 patterns is a consistent effect it should be enhanced when the level of this stimulus is increased. However, the conditions $\delta = 4$ and $\delta = 6$ do not support this hypothesis.

Under the conditions $\delta = 4$ and $\delta = 6$, the learning period was shorter and the subjects reached a steady-state performance after only about five sessions of 150 trials. The pooled data for two subjects over the last 900 trials indicates an increased detection ability over the $\delta = 2$

condition.

		STIMULUS		
		1	2	3
RESPONSE	1	.251	.040	.003
	2	.071	.236	.033
	3	.011	.057	.297

Pooled Data 900 Trials

$\delta = 4$

Detection Probability=.784

		STIMULUS		
		1	2	3
RESPONSE	1	.300	.070	0
	2	.033	.223	.020
	3	0	.040	.313

Pooled Data 900 Trials

$\delta = 6$

Detection Probability=.836

An answer to the second question posed in section 4.1 is now possible. When feedback is provided a human can learn to detect patterns which differ by as little as two second-order counts at a level of about 58% correct classification. When there is a difference between patterns of six second-order counts, about 84% of the patterns are classified correctly. In any of these cases guessing would account for only about 33% of the correct classifications.

Furthermore, although some subjects favored one type of information over the others in the early phases, this effect is based to a large extent on initial response bias, and diminishes after training. Such an effect becomes almost non-existent when the difference between patterns is large. This indicates that there is no large, consistent favoritism of any one type of second-order information after the subjects are well trained. Subjects can learn to use all types of second order information equally.

4.4 Discussion and Limitations of Results of Experiment 1

As mentioned earlier, in section 4.2, an optimum decision strategy for the class of patterns used in Experiment 1 involves simply counting the number of 11 sequences occurring in the target column and comparing this number to the proper decision thresholds, T_1 and T_2, located between the impulses of the density function of N_{11}. The optimum decision strategy is, in this case, 100% correct and, as such, a meaningful comparison with the subjects' performance is not possible. Also, because of the special nature of the patterns, there are very little data on which to base a measurement of the operator's psychometric function, i.e., the probability of a particular decison versus N_{11} for the target column. One may, however, hypothesize as to the form of the human psychometric function, and determine whether this hypothesis fits the data well or not.

Earlier work by Brazeal (2) and Glorioso (4) showed that a model of the human operator (in a detection task with first-order information) as an ideal detector with an inherent Gaussian distributed noise source fit the data very well. Using the same model in the present study results in a model detector which counts the number of 11's in the target column, adds a random number, $\Omega(\mu,\sigma^2)$-due to operator noise- and compares the sum to the decision thresholds. This model is depicted in Figure 4.3, and the associated probability densities are shown in Figure 4.4. Glorioso (4) found that for a four-choice decision human operators set decision thresholds very near the optimum values. By determining the value of the operator's decision thresholds and the variance of the operator noise (the mean is taken as zero when decision thresholds are allowed to vary) it is possible to fit a model to the operators' S-R

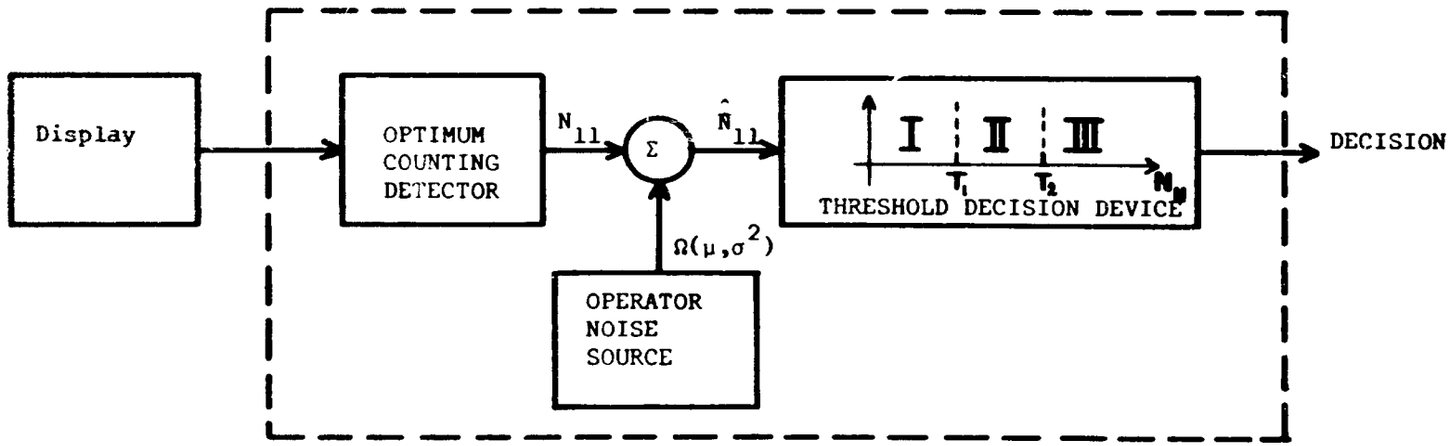


Figure 4.3

Model of Operator As An Optimum Detector with Internal Noise Source

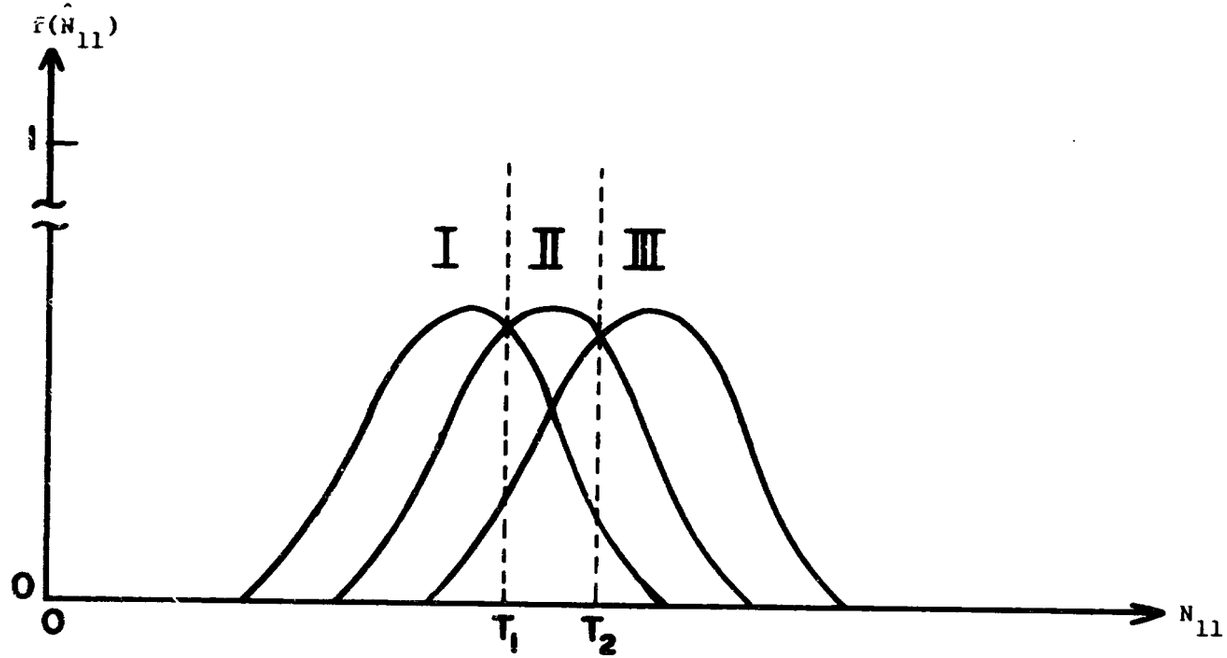


Figure 4.4

Statistics of Operator's Estimate, $\hat{N}_{11}$, of N_{11} in Target Column, and Decision Boundaries

matrices very closely. Using the values of thresholds and variance which closely fit the operators' performance, it was possible to obtain the modeled S-R and difference matrices of Figure 4.5. The elements, D_{ij}, of the difference matrices are given by,

$$D_{ij} = (M_{ij} - O_{ij}) \times 150$$

where M_{ij} is the element in the i^{th} row and j^{th} column of the model's S-R matrix and O_{ij} is the corresponding element of the operator's actual S-R matrix. The close agreement shown by such small values in the difference matrices is encouraging and lends support to the hypothesis that the operator can be modeled as an ideal detector with an additive Gaussian distributed noise source. Operator decision thresholds were set very near the optimum values, which are located at the intersections of the density functions in Figure 4.4. It is interesting to note that the operator noise variance is roughly constant, or, at least, that there is no apparent systematic change in operator noise over a wide range of stimulus (N_{11}) intensity. Compare this result with the approximately linear relation between operator noise and stimulus variance for first order information reported by Brazeal (2). The result agrees in that here the stimulus variance is constant (actually zero),and operator noise variance is also constant. It differs from Brazeal's result, however, in the existence of an operator noise with zero stimulus variance. This may be interpreted as a "fundamental" operator noise to which is added a term related to stimulus variance. Fitting S-R matrices is not, however, a particularly accurate method of determining operator noise, and the next chapter discusses more exact measurements through the use of psychometric functions. The purpose of the present discussion is only to point out that, even with little data, the possibility of a Gaussian distributed operator noise source for second-order information

$$[m_{ij}]_{\delta=2} = \begin{vmatrix} .197 & .076 & .014 \\ .119 & .172 & .109 \\ .017 & .085 & .210 \end{vmatrix} \qquad [d_{ij}] = \begin{vmatrix} 0 & +1.65 & -.45 \\ +2.4 & +.45 & +.45 \\ -2.4 & -2.1 & 0 \end{vmatrix}$$

$$[m_{ij}]_{\delta=4} = \begin{vmatrix} .252 & .029 & 0 \\ .080 & .236 & .036 \\ .001 & .069 & .297 \end{vmatrix} \qquad [d_{ij}] = \begin{vmatrix} +.15 & -1.65 & -.45 \\ +1.35 & 0 & +.45 \\ -1.50 & +1.80 & 0 \end{vmatrix}$$

$$[m_{ij}]_{\delta=6} = \begin{vmatrix} .300 & .044 & 0 \\ .100 & .222 & .060 \\ 0 & .067 & .313 \end{vmatrix} \qquad [d_{ij}] = \begin{vmatrix} 0 & -3.90 & 0 \\ 0 & -.15 & 0 \\ 0 & +4.05 & 0 \end{vmatrix}$$

	$\delta=2$	$\delta=4$	$\delta=6$
T_1	19.47	18.35	18.2
T_2	22.32	22.6	23.1
σ	2.04	1.95	2.5

$$[d_{ij}] = [m_{ij}] - [o_{ij}]$$

Figure 4.5

Modeled S-R and Difference Matrices for δ = 2,4,6

is quite feasible.

4.5 Summary of Results of Experiment 1

Two questions emerge as a result of Experiment 1. First, the patterns used in this experiment were very special in that, although they appeared to be random, they actually fell into three non-overlapping classes, and were perfectly identifiable by the simple strategy of counting the number of second-order sequences in the target column. More interesting, and necessary for an investigation of deeper questions, is a study of the larger class of patterns which may be generated by some statistical process which has dependencies between consecutive output symbols, for example, a Markov process. In these cases the pattern classes may overlap. That is to say, any one particular pattern may be generated (with different probability) by various statistical processes. No detector will be infallible for this larger class of patterns, and a comparison between the human and statistically optimum detector becomes meaningful. Also it is possible by using such displays to determine the precise form of the operator's psychometric function, i.e., the parameters of the operator noise. The question of human performance with patterns generated by a Markov process is discussed in the next chapter.

Second, information in Experiment 1 was provided only through a difference in second-order sequences. It is interesting to know not only whether or not a human can use dependent information, to what degree, and in what way, but also how dependent information is related to independent information in terms of its ability to be perceived. Is there some level of independent information above which dependent

information ceases to be a factor in determining human detection capability? By combining various amounts of independent and dependent information, the question of the relative utility of each is answered in Chapter 6.

Chapter 5

Experiment 2-Markov Displays

5.1 Introduction to Experiment 2

In the last chapter human performance in a visual detection task with a set of very restrictive patterns was discussed. This class of patterns was sufficient to answer some basic questions about human information processing of higher-order information, however, the answers obtained raised other questions. To answer these questions requires the use of patterns generated by a statistical process with inter-symbol dependencies. In the present chapter Experiment 2 is discussed in an attempt to answer the following questions.

First, when using a set of patterns, each of which has the possibility of being generated by more than one statistical process, does the human perform better or worse than with the restrictive (non-overlapping) set of patterns used in Experiment 1? Consider the problem of classifying a pattern which may have been generated by one of two statistical processes with densities described by the envelopes shown in Figure 5.1. The first process is assumed to be a binary statistically indapendent process with $P(1)=P(0)= 1/2$. It can be readily shown (see Appendix B.4) that the number of 11 sequences in a target column of length 84 is binomially distributed with a mean of $NP(11)=84 \times 1/4 = 21$ and variance of $NP(11)Q(11)=15.75$ where $Q(11)=1 - P(11)$. Let the second process be a first order Markov process with the same first-order probabilities as the statistically independent process. However, set the conditional probability, $P(1/1)$, such that it is greater than $P(1)= 1/2$. In particular let $P(1/1) = 0.642$ in which case the number of 11 sequences

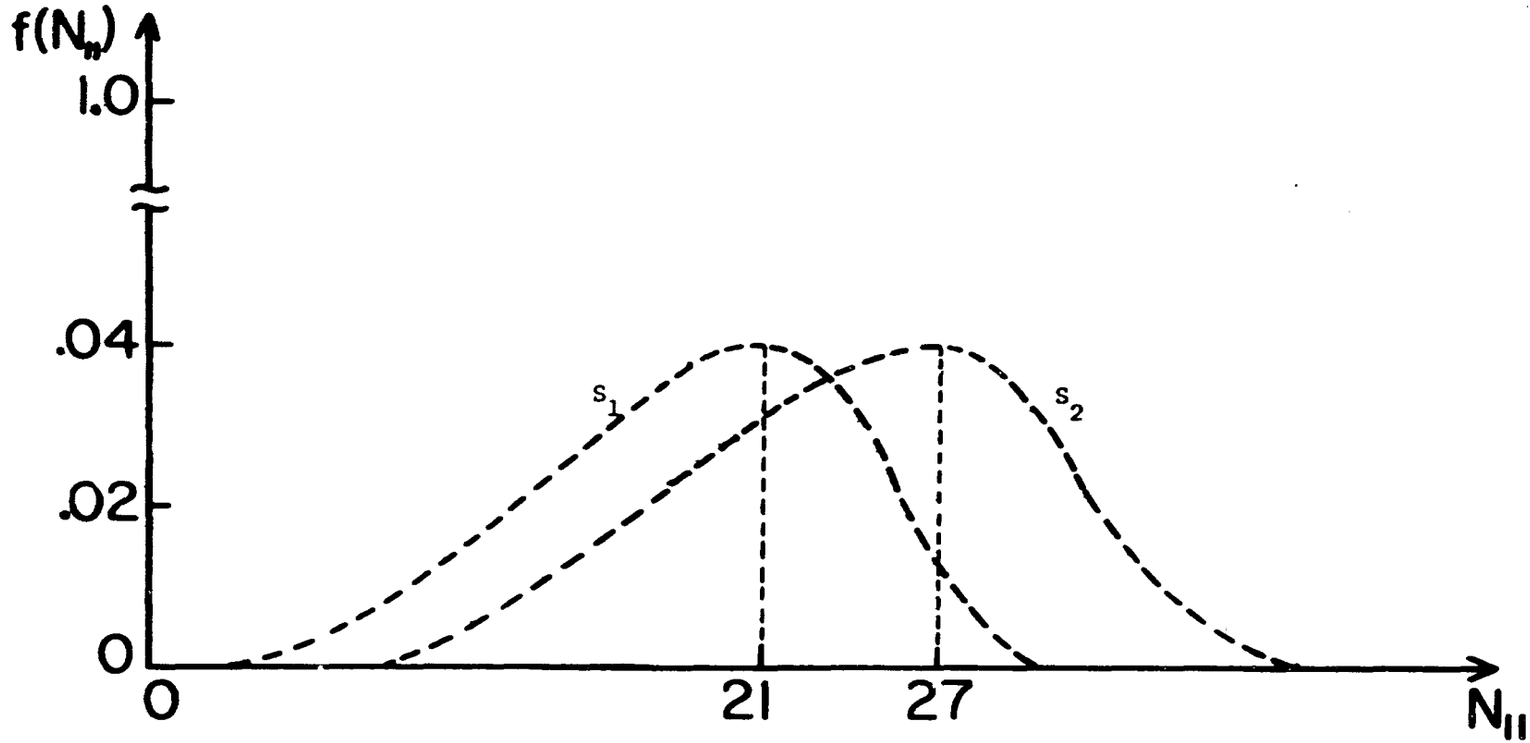


Figure 5.1

Envelope of Density Functions Related to Two Generating Sources

N_{11}, in the target column of length 84 is binomially distributed with mean of about 27 and variance of 17.75. It should be noted that the means of the two distributions shown in Figure 5.1 coincide with the impulses of the density function for N_{11} in Experiment 1 for stimulus-2 and stimulus-3, under the condition $\delta = 6$. With the sources shown, however, there is a non-zero variance in both distributions and some patterns will, therefore, be misclassified even by an optimum detector.

The second question which Experiment 2 answers is concerned with the form of the operator noise. Is the operator noise actually Gaussian distributed as the close fit obtained in Chapter 4 between the actual and modeled S-R matrices would suggest? Also, how does the operator perform compared to a statistically optimum detector?

In the last chapter, the parameter used by the optimum detector was the number of 11's in the target column. The optimum detector achieved 100% correct performance, and it was hypothesized that the human performed as an optimum detector corrupted by an internal operator noise, which was assumed to be Gaussian distributed. By using a first-order Markov process to generate the displays, it is possible to obtain a plot of the probability of the subject making a particular decision versus whatever decision parameter an optimum detector would use. For an optimum detector the decision strategy results in a sharp boundary at some decision threshold, T, as indicated by the solid line in Figure 5.2. All patterns with a/value of decision parameter, P, greater than T are put into one class, and the rest into another class. The human, however, cannot accurately determine the value of P for each pattern. Thus, his classification performance (see dotted line in Figure 5.2) in general

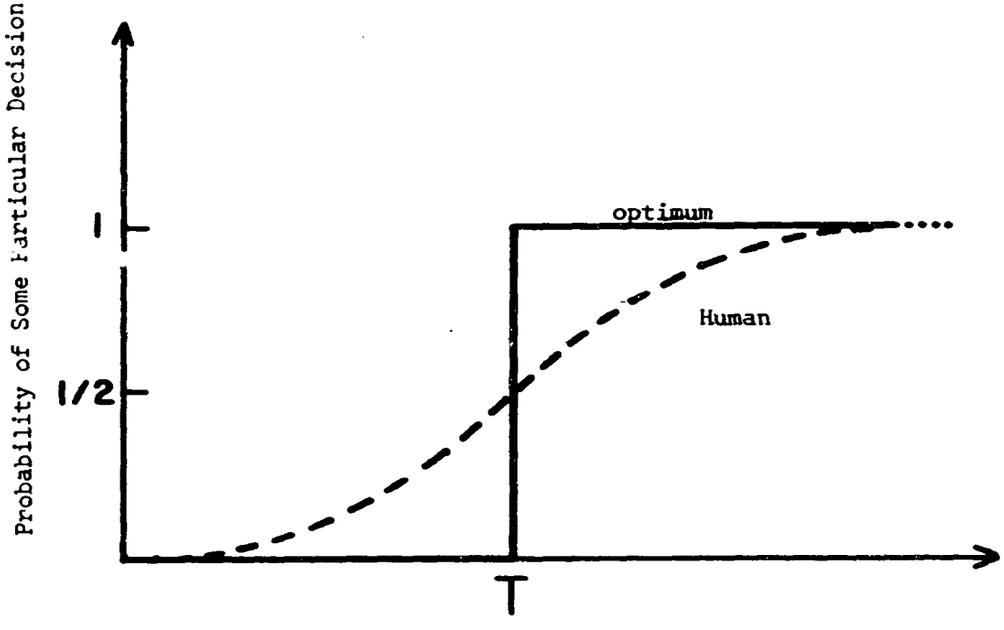


Figure 5.2

Optimum Detector and Typical Human Psychometric Functions

only approximates that of the ideal detector. This "psychometric function" characterizes the operator's use of the particular parameter as a cue in detection. If the resulting curve follows a cumulative Gaussian distribution the model is supported. In this chapter a measure of the mean and variance of the operator noise is obtained by such a method.

If the human psychometric function is actually Gaussian distributed, the standard deviation of the operator noise may be determined by taking one half the difference in parameter values, P, which correspond to probabilities of 0.16 and 0.84. A useful psychological measure of human sensitivity is the "just noticeable difference", or j.n.d., which may be defined as one half the amount of stimulus change necessary for a change in probability of classification of 0.5. From the psychometric function a j.n.d. is one half the change in P which corresponds to a change in probability from 0.25 to 0.75

5.2 Design of Experiment 2

Making use of the specific form of the optimum detector expressed by relation 3.4.2, with a basis set $\{N, N_1, N_{11}\}$, the optimum decision strategy for patterns which may be generated by either a first-order Markov process or a statistically independent process may be expressed as

5.2.1

$$84 \log P(0/0)+N_1 \log \frac{P(1/0)P(0/1)}{P^2(0/0)} +N_{11} \log \frac{P(0/0)P(1/1)}{P(0/1)\,P(1/0)} \gtrless 83 \log 1/2$$

: D_2 (Markov)

otherwise : D_1 (Statistically Independent)

This relation may be further simplified to read,

$$N_1 k_1 + N_{11} k_{11} \gtrless \log \frac{1}{2P(0/0)} : D_2 \text{ (Markov)}$$

$$\text{otherwise} \quad : D_1 \text{ (Statistically Independent)} \qquad 5.2.2$$

where $k_1 = \log \frac{P(1/0)P(0/1)}{P^2(0/0)}$

$$k_{11} = \log \frac{P(0/0)P(1/1)}{P(0/1)P(1/0)}$$

As pointed out in Chapter 3 and discussed in detail in Appendix B.2 , k_1 and k_{11} may not vary independently. What, then, is the form of the displays which may be presented in an experiment which uses a first-order Markov process and a statistically independent process? For the questions which are to be answered by Experiment 2 it is desirable to use a Markov process which results in the simplest decision strategy. From the weighted summation of equation 5.2.2 it is obvious that the case $k_1 = \pm k_{11} = k$ would be a desirable choice. However, as pointed out in Appendix B. 2 , the condition $k_1 = k_{11}$ is impossible, but $k_1 = -k_{11}$ is entirely feasible. If $k_1 = -k_{11} = k$ it is shown in Appendix B.3 that the Markov transition matrix must be double stochastic; this implies equal first-order probabilities, $P(0) = P(1) = 1/2$, and the Markov process is completely specified. Although it will not be verified until Chapter 6, one other reason for choosing $k_1 = -k_{11}$ is that this condition corresponds to what will later be called "purely dependent" information content in the display. This added condition is not necessary to answer the questions asked in the present chapter, but the proper choice at this point provides a bonus when combination of information is discussed later.

Thus for the present, only the limited case of $k_1=-k_{11}=k$ will be studied. Under this assumption the optimum decision rule involves counting the number of 1's and the number of 11's in the target column and comparing the difference to an appropriate threshold.

$$N_1-N_{11} \gtrless \frac{84}{k} \log \frac{1}{2P(0/0)} \quad : \quad D_1 \text{ (Statistically Independent)}$$

$$\text{otherwise} \quad : \quad D_2 \text{ (Markov)} \qquad 5.2.3$$

$$\text{Note} \quad : \quad k < 0$$

For the case in which the observed sequence is much longer than the order of the Markov process ($r << t$), N_1-N_{11} is approximately equal to N_{01} or N_{10}, following the reasoning used in section 3.3.

With the optimum decision parameter, $N_1-N_{11} \gtrless N_{10}$, specified, the conditions on P(1/1) must be determined. Since it was found in Experiment 1 that the subject did not favor either an increase or a decrease in N_{11} over the opposite situation, the "one sided" case in which P(1/1) > P(1) was used in Experiment 2. Two conditions, as outlined in Table 5.1, were studied. The means of the N_{11} distributions governing the generation of patterns were set to correspond to stimulus-2 and stimulus-3 patterns of Experiment 1, for the two conditions $\delta = 2$ and $\delta = 6$.

The same three subjects participated in all display conditions in Experiment 2. Subject A was also subject A in both phases of Experiment 1, while two additional subjects, E and F, both undergraduate engineering students paid for their services, participated in this experiment. Each session consisted of 100 trials rather than 150 used in Experiment 1 in order to reduce any undesirable effects due to fatigue. On each trial the subject was required to make one of two decisions which were indicated by pressing one of two buttons located in front of him. The

CONDITION		P(1/1)=P(0/0)	P(1/0)=P(0/1)	$E\{N_{11}\}$	Variance (N_{11})	Variance (N_{10})
A	S_1	.5	.5	21	15.75	15.75
	S_2	.642	.358	27	18.3	12.35
B	S_1	.5	.5	21	15.75	15.75
	S_2	.543	.457	23	16.65	14.8

P(0)=P(1)= 1/2

Subjects: A,E,F

} For all conditions.

TABLE 5.1 CONDITIONS OF EXPERIMENT 2

possible decisions were:

D_1: Display generated by source -1 - a statistically independent process with P(1)=P(0)= 1/2.

D_2: Display generated by source-2 - a Markov process with statistics known to the subject before running.

Before the first session the statistics related to the generation of displays by each source were explained to the subjects and after each trial knowledge of results was provided in the form of an "H" for "hit" or an "M" for "miss" (see Chapter 2). The subjects were told to work as quickly as possible without diminishing confidence in their decisions. After each decision, the computer determined the first and second-order sequence counts in the target column, and typed out the following data:

-subject's decision,

-correct decision,

-N_1, N_0 in the target column,

-N_{11}, N_{00}, N_{10}, N_{01} in the target column,

-subject's decision time

At the end of each session the subject's S-R matrix was outputed, and the subject was told how well he had performed. All of the data for each session were also recorded on paper tape for further processing.

5.3 Results of Experiment 2

The main goal of Experiment 2 was the determination of the form of the human psychometric function, and, thus, the form of the operator noise. As such, only data representative of the subjects' steady-state performance, such as those obtained from the later sessions, were retained

for further processing. The data from these later sessions, consisting of an average of 600 trials per subject, were processed by a special computer program which extracted the information necessary to plot the psychometric function.

The program first calculated the value of the parameter N_1-N_{11} in the target column for each trial. Recall from section 5.1 that N_1-N_{11} is the parameter used by the optimum detector in making a decision. Based on the value of this parameter, the remaining information from each trial was categorized and summed over all trials. This procedure provided the following measures:

-number of times a pattern appeared for each value of N_1-N_{11},

-number of times a Markov pattern appeared for each value of N_1-N_{11},

-number of times decision-2 (Markov display) was made by the operator for each value of N_1-N_{11}.

From these processed data the subject's psychometric function was obtained. Also, the overall S-R matrices for both the subjects and the optimum detector were calculated.

Figure 5.3 shows the S-R matrices for both the pooled subject data and the optimum detector under both experimental conditions, A (strong dependency) and B (weak dependency).

In Chapter 4 the question was raised of whether or not the subject would perform better with the overlapping set of patterns used in this experiment. The stimuli used in the present experiment correspond with respect to means of the probability density function of N_{11} to the stimulus-2 and stimulus-3 conditions of Experiment 1. However, there is no counterpart in the present experiment to the stimulus-1 condition

		STIMULUS	
		1	2
Response	1	.40	.10
	2	.09	.41

Subjects, P(D)=.81

		STIMULUS	
		1	2
Response	1	.44	.06
	2	.06	.44

Optimum, P(D)=.88

Condition A

		STIMULUS	
		1	2
Response	1	.31	.18
	2	.21	.30

Subjects, P(D)=.61

		STIMULUS	
		1	2
Response	1	.42	.12
	2	.10	.36

Optimum, P(D)=.78

Condition B

Figure 5.3

S-R Matrices for Pooled Subject

Data and Optimum Detector

of Experiment 1. As such, it might be argued that in comparing performance in these two tasks, classification of stimulus-2 as stimulus-1 (in Experiment 1) should actually be considered as correct classification of stimulus-2. Making such an assumption, the modified probability of detection (correct classification) assuming only the presence of stimulus-2 and stimulus-3 patterns, is 0.688 for $\delta = 2$ and 0.91 for $\delta = 6$. Comparing these values to the corresponding probabilities of detection in Experiment 2 of 0.61 and 0.81 indicates that the patterns from the overlapping set used in Experiment 2 are consistently more difficult to classify than those chosen from the restrictive, non-overlapping set used in Experiment 1.

It is not meaningful to compare the optimum detector's performance in Experiment 1 to that shown in Figure 5.3 for Experiment 2 since the former achieved 100% correct performance. However, comparing the subjects' performance to that of the optimum detector demonstrates, as expected, the superior ability of the optimum detector. Notice for strong dependencies, however, that the subjects' 0.81 detection probability compares quite favorably to 0.88 obtained by the optimum detector.

Figure 5.4 is a plot of the psychometric function for the three subjects participating in Experiment 2 under the condition of strong inter-symbol dependency. The abscissa of this figure is a normal probability scale, thus, a cumulative Gaussian distribution plots a s a straight line. Notice that a particular distance at the extremities represents much less change in probability than an equal distance near

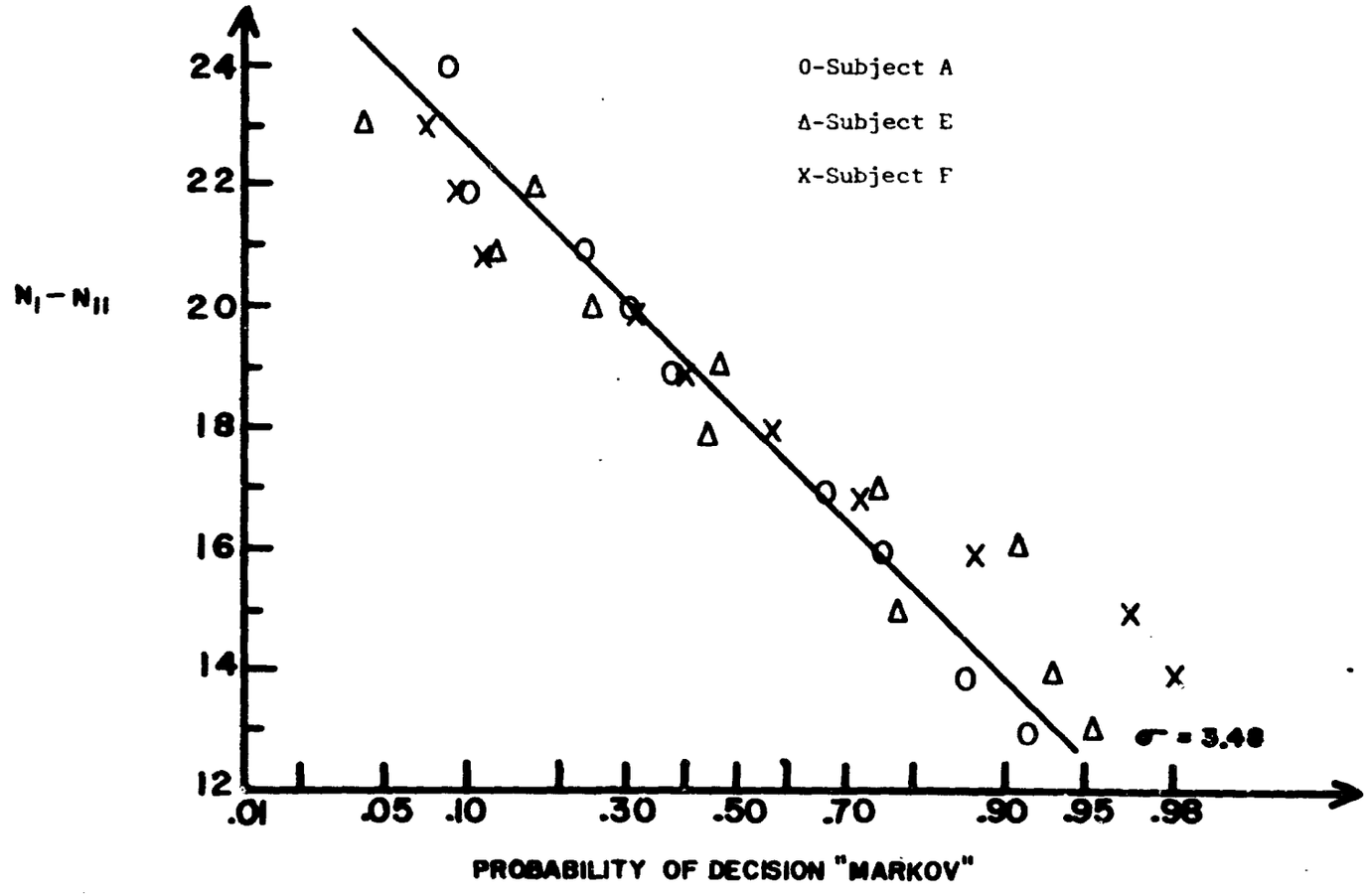


Figure 5.4

Average Subject Psychometric Function, Condition A, Experiment 2

	Mean	Standard Deviation	Mean Square Error
Subject A	18.03	3.56	.0011
Subject E	18.34	2.75	.0046
Subject F	18.54	2.44	.0067

Table 5.2

Best Fit Gaussian Approximation of

Operators' Psychometric Function, Condition A, Experiment 2

the center. A computer program was used to find a best fit (minimum mean square error) Gaussian distributed approximation to the data points for each subject. The results are shown in Table 5.2. Since the mean of these distributions were nearly equal, the variances were averaged to obtain an overall best fit model of the pooled subject psychometric function. This is shown in Figure 5.4 as the straight line with mean of 18.2 and standard deviation of 3.48. It should be noted that the mean is extremely close to the optimum decision threshold of 18.4 calculated from equation 5.2.3.

Under the condition of weak inter-symbol dependency, condition B, the subject's data points, Figure 5.5 were not very consistent. Best fit Gaussian distributed models of each subject's psychometric function are shown in Table 5.3. The mean of 19.92 used by Subject A was very close to the optimum decision threshold of 19.85, however, the other two subjects deviated considerably. By adjusting the subjects' data points so that the resulting means coincided with the optimum decision threshold, Figure 5.6 was obtained. A "best fit" Gaussian distributed model of the pooled subject psychometric function is shown by the straight line with mean of 19.85 and standard deviation of 5.22. However, this model is strongly biased by the extreme variance shown by Subject F. Deleting Subject F's data points results in the model with standard deviation of 3.64. It is obvious that a precise measure of the variance of the operator's psychometric function under the condition of weak dependency is not possible, however, a value between 3.6 and 5.2 seems appropriate. Also, a value of j.n.d. of from 2.5 to 3 second-order counts is indicated by the psychometric functions.

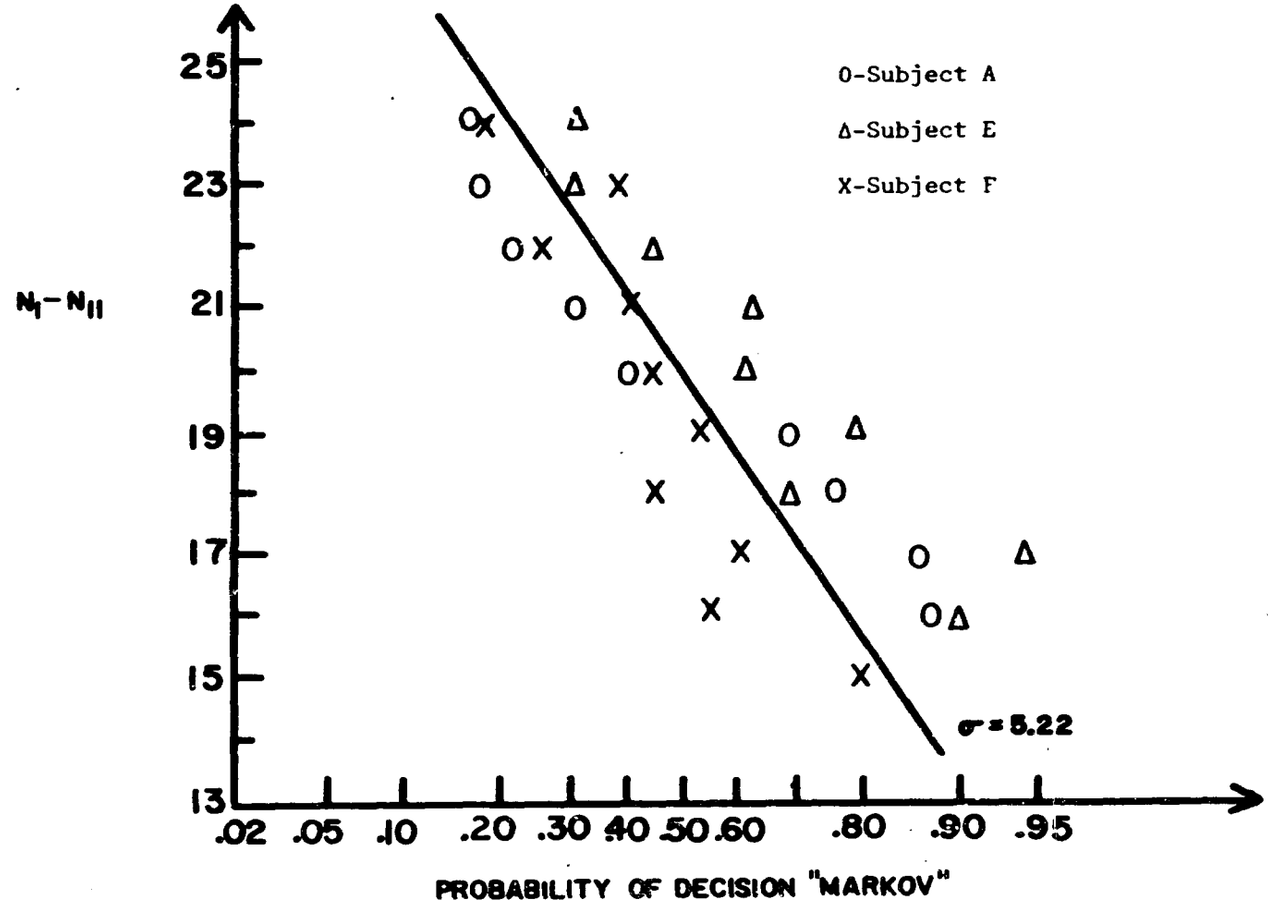


Figure 5.5

Average Subject Psychometric Function, Condition B, Experiment 2

	Mean	Standard Deviation	Mean Square Error
Subject A	19.92	3.076	.0028
Subject E	21.60	4.204	.0039
Subject F	18.41	8.409	.0036

Table 5.3

Best Fit Gaussian Approximation of Operators'

Psychometric Function, Condition B, Experiment 2

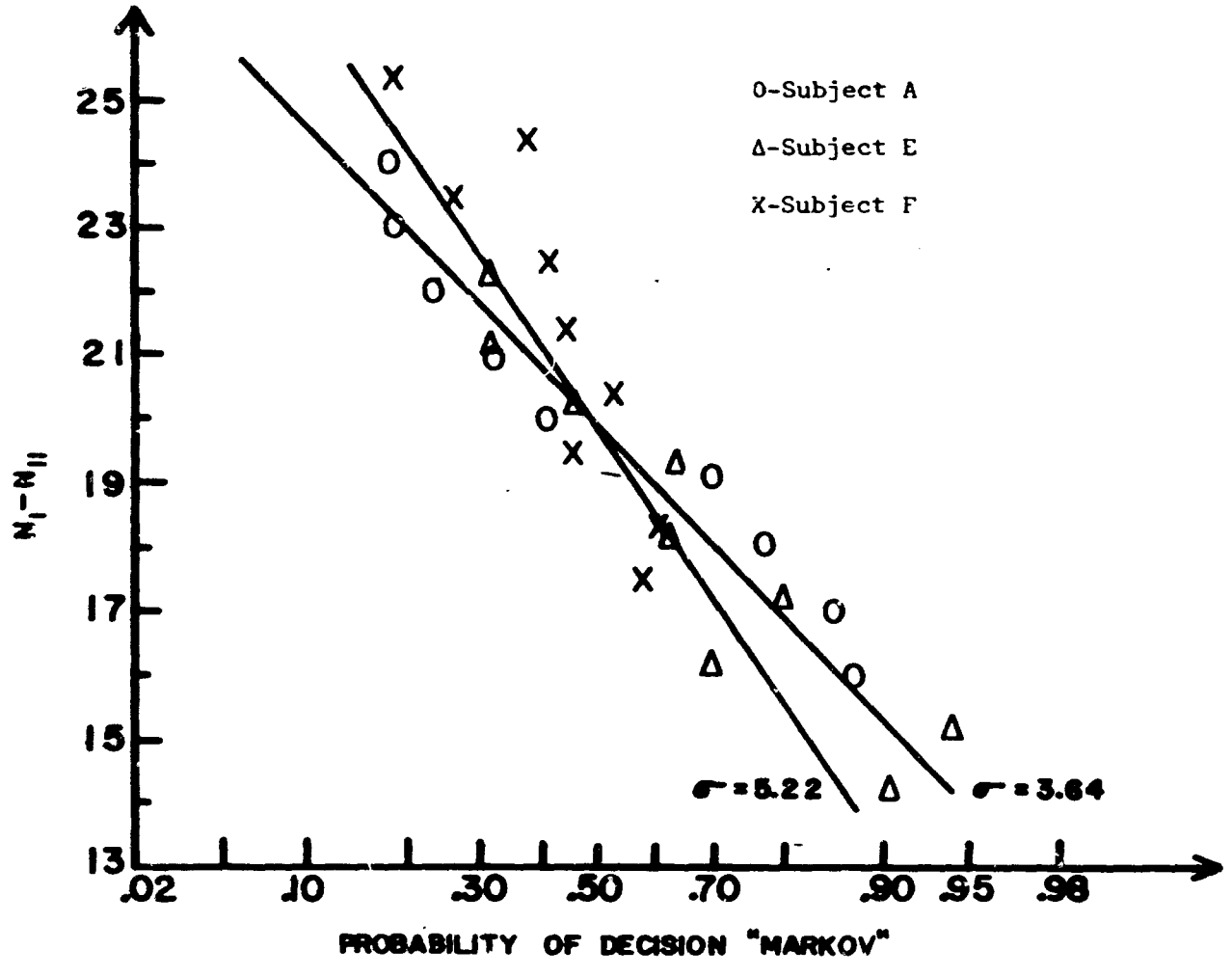


Figure 5.6

Subjects' Modified Psychometric Function, Condition B, Experiment 2

5.4 Discussion and Summary of Results of Experiment 2

This section provides an interpretation of the experimental findings of Experiment 2, and a summary of results.

It is apparent that with strong dependencies the subject may set a threshold very near to the optimum decision threshold, and that he appears to operate with an internal operator noise which is Gaussian distributed with a standard deviation of about 3.5. However, when there are only weak dependencies (P(1/1)= .543 in this case), the subject does not set his decision threshold as precisely. Nevertheless, it is still set near the optimum value. With weak dependencies, the value of operator noise variance varies considerably between subjects, but is consistently larger than for strong dependencies. Calculation of the precise relation between operator noise variance and stimulus variance is not possible with the data available. However, it is clear that operator noise variance and variance of the cue used by the operator (N_{10} in this case), are directly related, as Brazeal found for first-order information. As an approximation, the linear relation found by Brazeal results in:

$$\sigma^2 = k^2 \, VAR(N_{10})$$

with a vaule of k^2 of about 1. For first order information, Brazeal found that a value of $k^2 = 1/2$ described the subjects' performance well. It is clear that operator noise variance is greater by a factor of about 2 when the cue used for detection is a second-order rather than a first-order parameter.

In summary, Experiment 2 has pointed out the following factors

related to human information processing with dependent, statistical, visual information:

1. In a two-choice decision task, humans can learn to use statistical information related to the inter-symbol dependencies of the source. Correct classification performance rises from about 60% when the pattern classes are separated by dependencies of about 0.043 (i. e., P(1/1) of 0.5 and 0.543) to a level of about 80% with 0.143 separation between dependencies of the two pattern classes.
2. Subjects learn to set near optimum decision boundaries, indicating that the mean of the operator noise is near zero. The decision thresholds are set more accurately when dependent information is strong than when it is weak.
3. Operator noise variance in a task involving dependent information is about twice as great as in a task using first-order information as the cue. The variance ranges from about 12 for patterns with strong dependencies to roughly 20 for patterns with weak dependencies.

Chapter 6

Experiment 3 - Combination of Information

6.1 Introduction to Experiment 3

In an attempt to determine to what extent various cues are used in visual pattern detection, it was suggested in Chapter 4 that subjects be presented with patterns containing various amounts of independent information, which is related to the individual symbol probabilities, and dependent information, which arises from the joint probability structure of the underlying process. Although we know the form and magnitude of operator noise for purely independent and purely dependent information, there are no data which pertain to the operator's relative use of each type of information when they are presented simultaneously. By measuring the probability of detection under various conditions of independent and dependent information, the answers to the following questions might be obtained. How much dependent (second-order) information is equivalent to a particular amount of independent (first-order) information? How does the human performance compare to an optimum detector when more than one type of information is present? When "equal amounts" of information on both levels are presented, which is used the most? Over what range is independent information superior to dependent information in visual pattern detection?

Before an experiment can be designed to answer these questions, it is necessary to give a more precise meaning to the term "amount of information", and how it is related to the visual displays used in this paper.

Consider, once again, the problem proposed by Figure 3.1. One of two sources is chosen at random to produce outputs, on the basis of

which an observer is to determine which source is the generating source. It is assumed that the output symbol sets are identical, and that the sources differ only in their underlying probability structure. At first, assume that both sources are governed by binary, statistically independent processes. Furthermore, assume that $P(1) \geq P(0)$ for both sources; this implies $P(1) \geq 1/2$.

Under these conditions, the entropy (10), $H(S_i)$, of each (i^{th}) source lies between one and zero and decreases monotonically with increasing $P(1)$.

$$H(S_i) = -[P_i(0) \log P_i(0) + P_i(1) \log P_i(1)] \quad 6.1.1$$

A measure of the "dissimilarity", U, of the two sources is proposed as,

$$U = |H(S_1) - H(S_2)| \quad 6.1.2$$

If the sources are very dissimilar (U is high) their probability structures (just P(1) in this case) must differ greatly. Note that $0 \leq U \leq 1$.

Assume, now, that one source, S_1 say, always has $P(1) = 1/2$, and $H(S_1) = 1$. Since $H(S_2) \leq 1$, the dissimilarity is,

$$U = 1 - H(S_2) \quad 6.1.3$$

and represents a measure of how greatly the probability structure of S_2 differs from that of source S_1, or pure chance. What, now, if S_2 (henceforth called simply S) is allowed to be governed by a first-order Markov process?

The total dissimilarity, U, is composed of two parts, one part due to the independent information (i.e., the first-order probability structure of the Markov source), and the other part arising from inter-symbol dependencies. Call these the independent dissimilarity, U_I, and the dependent dissimilarity, U_D, respectively. Thus $U = U_I + U_D$. To obtain a quantitative measure of each component, consider a source

$\bar{S}$, called the adjoint source (10), which has P(1) and P(0) equal to the first-order probabilities of S. However, let there be no inter-symbol dependencies in $\bar{S}$, i.e., P(1/1) = P(1/0) = P(1) and P(0/0) = P(0/1) = P(0). Furthermore, let S^2 be a source which has output symbols, σ_i, composed of pairs of output symbols of S, and let $\bar{S^2}$ be the adjoint of S^2. Thus the probability $P(\sigma_i)$ of each output symbol from S^2 is equal to the probability of sequences of length two from S. It is shown in Appendix B.5 that the entropy of a Markov source is,

$$H(S_m) = H(\bar{S^2}) - H(\bar{S}) \qquad 6.1.4$$

For example, assume that S is a binary first-order Markov process with,

$$\begin{aligned} &P(0/0) = P(1/1) = 0.7 \\ &P(0/1) = P(1/0) = 0.3 \\ &P(0) = P(1) = 0.5 \end{aligned} \qquad 6.1.5$$

$\bar{S}$ is a statistically independent source with,

$$P(0) = P(1) = P(0/0) = P(1/0) = P(0/1) = P(1/1) = 0.5, \qquad 6.1.6$$

and S^2 has an output symbol set, and symbol probabilities, of:

$$\begin{aligned} &\sigma_1 = 00 && P(\sigma_1) = 0.35 \\ &\sigma_2 = 01 && P(\sigma_2) = 0.15 \\ &\sigma_3 = 10 && P(\sigma_3) = 0.15 \\ &\sigma_4 = 11 && P(\sigma_4) = 0.35 \end{aligned} \qquad 6.1.7$$

Also, $\bar{S^2}$ is a statistically independent process with first-order symbol probabilities the same as those of 6.1.7. From the above probabilities, the entropy of the Markov source, S, may be calculated.

$$\begin{aligned} H(S_m) &= H(\bar{S^2}) - H(\bar{S}) \\ &= (2)(.35 \log_2 1/.35) + (2)(.15 \log_2 1/.15) \\ &\qquad - (2)(.5 \log_2 1/.5) \\ &= 1.8813 - 1 = .8813 \end{aligned} \qquad 6.1.8$$

Returning, now, to the problem of determining the individual components of dissimilarity, assume that $\bar{S}$ were the actual generating source. There would be no dependent information available to contribute to dissimilarity, and the independent component of dissimilarity is,

$$U_I = 1 - H(\bar{S}) \qquad 6.1.9$$

Since the total dissimilarity, U, is the sum of the component dissimilarities, the additional amount of dependent dissimilarity arising from actually using S, a Markov source, instead of $\bar{S}$, the adjoint source. must be,

$$\begin{aligned} U_D &= U - U_I \\ &= [1 - H(S)] - [1 - H(\bar{S})] \\ &= H(\bar{S}) - H(S) \\ &= 2H(\bar{S}) - H(\bar{S}^2) \end{aligned} \qquad 6.1.10$$

Equations 6.1.9 and 6.1.10, along with tables of entropy for various sources, permits the design of a unique source for any specified component dissimilarities.

As a measure of the "dependency" in patterns arising from the use of sources described above, the following ratio is proposed, and is used in the remaining development.

$$D = \frac{U_D}{U_D + U_I} \quad , U = U_D + U_I \qquad 6.1.11$$

Dependency, D, ranges from zero, for a statistically independent process ($U_D = 0$), to +1 for a Markov process with $P(1) = P(0) = 1/2$ ($U_I = 0$). A D measure of 1/2 indicates that equal amounts of independent and dependent components of dissimilarity are present.

Using the above definitions of "dissimilarity" and "dependency", Experiment 3 was designed to answer the questions posed earlier.

6.2 Design of Experiment 3

By using either a statistically independent process with $P(1) = P(0) = 1/2$, or a first-order Markov process, patterns were generated which contained the same total amount of dissimilarity as the patterns used in Experiment 2, however, they possessed varying amounts of component dissimilarities, U_I and U_D. Table 6.1 summarizes the experimental conditions used in Experiment 3. There were two amounts of total dissimilarity in the displays, $U = .007$ and $U = .06$, with three levels of dependency, $D = 1/3$, $1/2$, and $2/3$. Data from Experiment 2 and interpolation from the results of Brazeal (2) fill in the cases of $D = +1$ and $D = 0$, respectively.

A new subject, G, was added to those who had participated in the past experiments. SubjectsA and B participated in all conditions with $U = .007$, while Subjects F and G ran all conditions of the experiment with $U = .06$. All subjects were required to make one of two decisions, Markov or Independent display, as in Experiment 2, on each trial. There were 100 trials per session. As before, feedback of knowledge of results was provided immediately after each decision, and the subjects were informed of their overall level of performance after each session. Each subject participated in an average of seven sessions for each condition. Subject detection probability rose rapidly in early sessions and leveled off to a value which varied less than 7% over the last three sessions. Because of this steady performance, and the fact that all subjects except G had

Condition	Dissimilarity, U	Dependency, D	P(0/0)	P(0/1)	P(1/0)	P(1/1)	P(0)	P(1)
A	.06	1/3	.495	.315	.505	.685	.384	.616
		1/2	.525	.315	.475	.685	.399	.601
		2/3	.558	.320	.442	.680	.416	.584
B	.007	1/3	.490	.435	.510	.565	.460	.540
		1/2	.50	.435	.50	.565	.465	.535
		2/3	.513	.435	.487	.565	.472	.528

Table 6.1

Conditions of Experiment 3

participated in other experiments, the data from the subjects' last three sessions were taken to be a measure of the subjects' steady-state performance.

6.3 Results and Discussion of Results of Experiment 3

With total "dissimilarity", U, of 0.06, average subject detection (correct classification) probability, P(D), was about 0.85. Figure 6.1 indicates the change in P(D) with dependency, D, for the two subjects participating in Experiment 3. Data points for the D = 0 and D = 1 conditions are taken from other work, as mentioned earlier, and there are no data for Subject F at D = 1/2. The three lines in Figure 6.1 compare the performance of an optimum detector (for Markov sources), the average of the two subjects, and a first-order detector (one which uses only independent information).

It is clear that the subjects' performance rose when less dependent (more independent) information was presented; however, the change was only on the order of 10%. The subjects seem to perform very much like a poor Markov optimum detector. This result agrees with the results from the psychometric functions obtained in Experiment 2, although only dependent information was used there. For dependent dissimilarity U_D greater than 50% of the total, subjects outperformed the first-order detector.

When total dissimilarity was only 0.007, Subjects A and B performed with a probability of detection of about 0.60. Average subject performance shown in Figure 6.2 indicates that there is very little change in P(D) over the complete range of dependency. An optimum Markov detector achieves about 68% correct decisions, and, again, subjects perform roughly 10% worse than the optimum detector. The

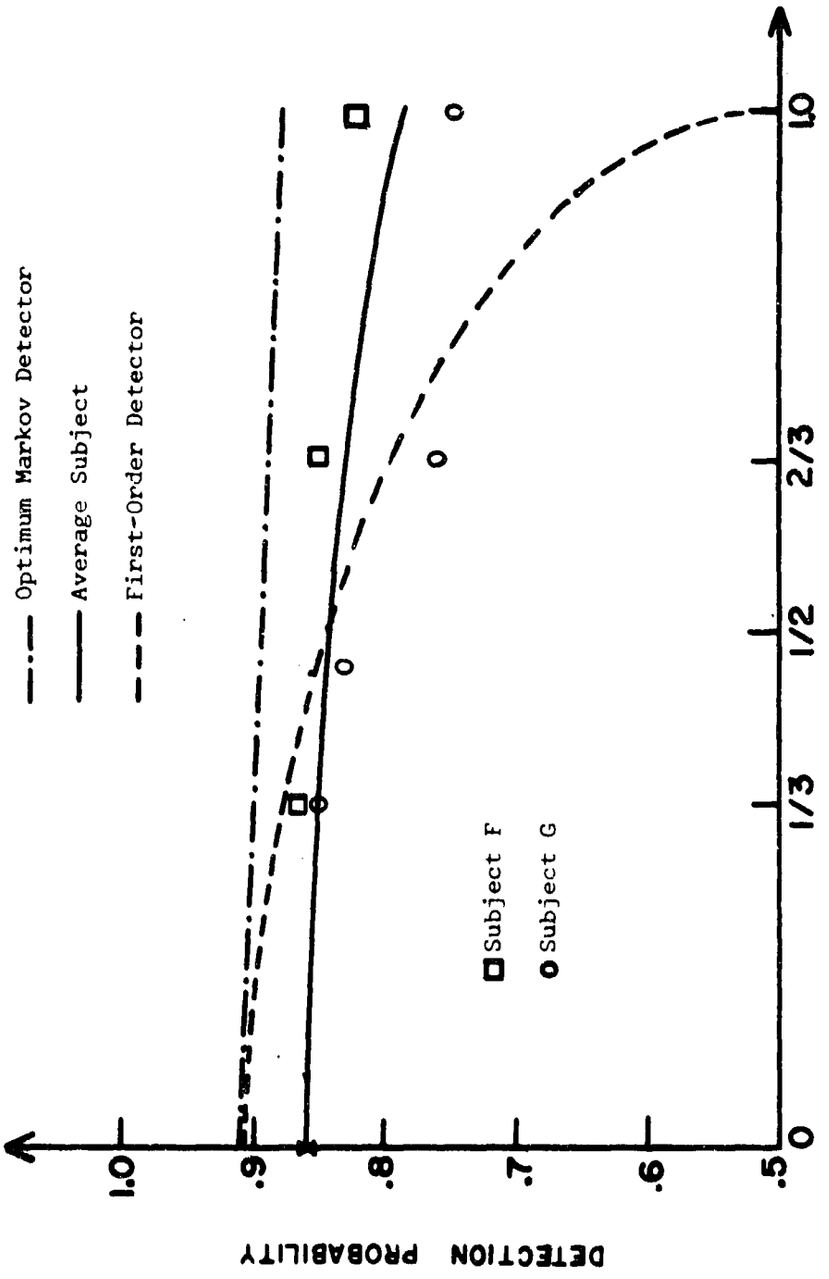


Figure 6.1

Probability of Detection versus Dependency, Condition A, Experiment 3

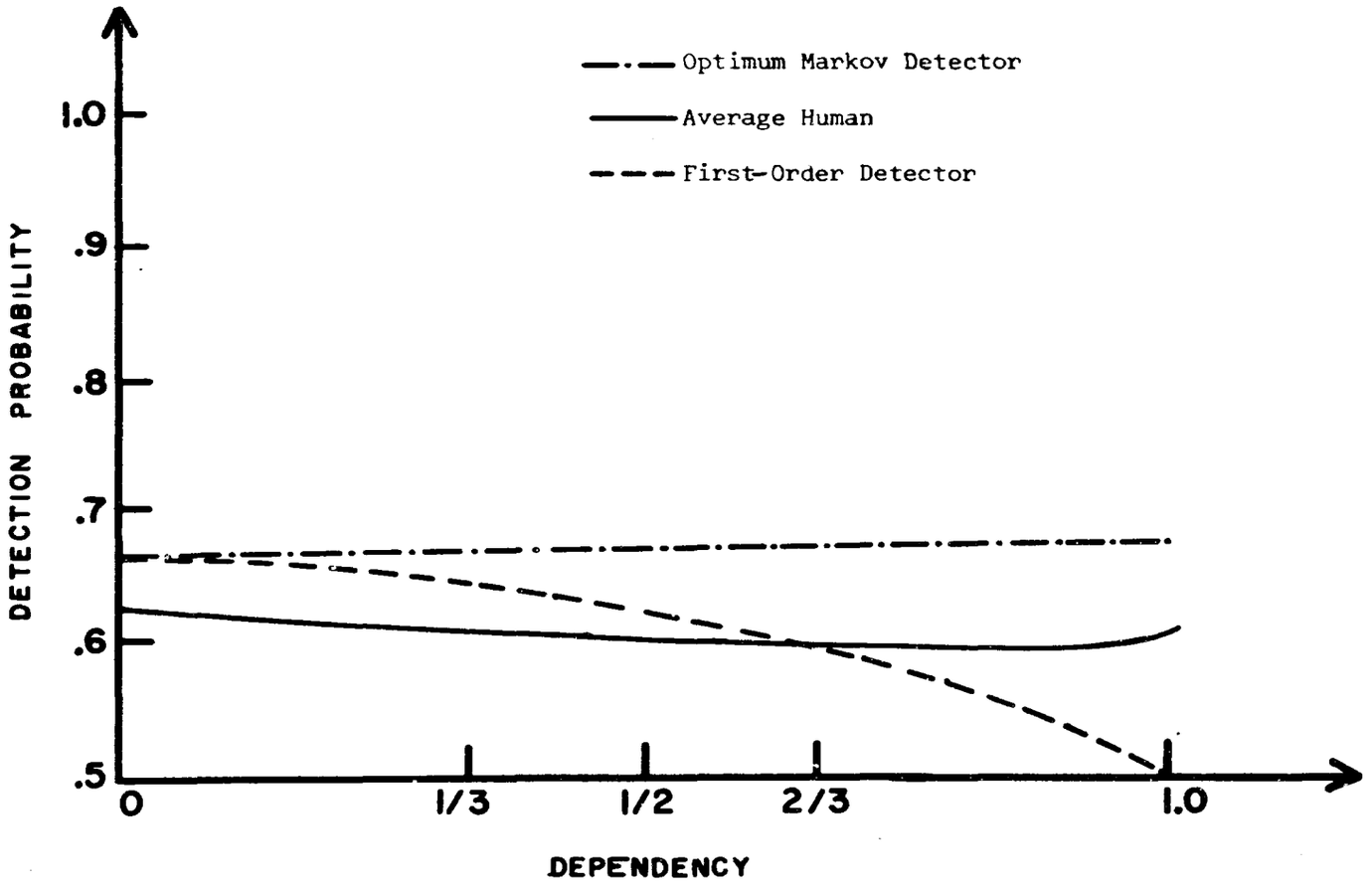


Figure 6.2

Probability of Detection versus Dependency, Condition B, Experiment 3.

first-order detector is superior to the subjects until dependencies make up about 2/3 of the total dissimilarity.

From this experiment we see that, although operator noise is greater for dependent information than for independent information, subject performance suffers by using greater amounts of dependent information only when the total information is high. At low levels performance is roughly constant, irrespective of the level of dependency. In both cases studied, the subjects' detection probability followed the form of the optimum detector, and not the first-order detector. Apparently, independent information can be extracted more accurately, but its presence never causes the trained subject to ignore the available dependent information.

Also, we see that if the dependency, D, of the patterns is less than 1/2, implementation of the simple first-order detector, which only counts the number of 1's in the target column and compares this to a threshold, provides performance superior to that of the human operator who uses the dependent information as well. However, when the dependent dissimilarity is high, the performance of the first-order detector deteriorates rapidly.

Chapter 7

Conclusions

7.1 Objectives and Method

This thesis has attempted to provide a better understanding of the effects of inter-symbol dependencies on human visual information processing ability. Three general experiments have provided the answers to the following questions.

1. Is the human operator inherently sensitive to information provided by inter-symbol dependencies? If so, within what range?
2. With extended practice, what level of performance can the human operator achieve in a visual detection task involving dependent statistical information?
3. Does the model of the human operator as an optimum detector corrupted by an internal noise source hold for a task involving dependent information? What is the form of the operator noise?
4. When presented with patterns containing both dependent and independent statistical information, does the human operator use one component of the information to a greater extent than the other?

The first experiment determined the range of human sensitivity to dependent information. Experiment 2 proceeded to determine the form of the human operator noise through the use of an ideal detector and experimentally derived psychometric functions. Experiment 3 provided definitions of "dissimilarity" and "dependency" of patterns generated by either a statistically independent or a Markov process. It then went on to discuss the relative usefulness of independent and dependent information.

7.2 Summary of Results

From the three experiments conducted, the following results were obtained:

1. It was found that the human operator possesses an inherent ability to recognize differences in patterns on the basis of second-order sequence counts only, provided that the patterns are separated by at least five or six second order counts.
2. With extended practice in a three choice decision task with patterns from non-overlapping classes, the human operator can learn to consistently classify patterns which differ by only two second-order counts at about a 60% level.
3. Classification of patterns drawn from overlapping classes used in Experiment 2 was consistently more difficult than classification of patterns from the non-overlapping classes used in Experiment 1 over a range of separation of N_{11} (or its mean) of from 2 to 6 counts.
4. Operator noise in a pattern detection task with dependent statistical patterns was found to be approximately Gaussian distributed with near zero mean and a standard deviation of from 3.5 to 5. The variance of the operator noise is roughly twice the variance associated with operator noise in a similar task using statistically independent visual information.
5. Operator performance, as measured by probability of detection, is better for independent information than for dependent information when the overall level of information is high, specifically $U = 0.06$. At low levels, $U = 0.007$, performance is

nearly constant, irrespective of the form of the information.

6. No point was found at which operators overlooked the presence of dependent information. Even when independent information made up a large portion of the total amount, operators made use of whatever dependent information was present.

7. For a level of dependency less than about 1/2 a simple first-order detector is capable of outperforming the human operator; however, the performance of this simple detector falls off rapidly as dependency increases above 1/2.

Appendix A

Computer generation of Markov sequences

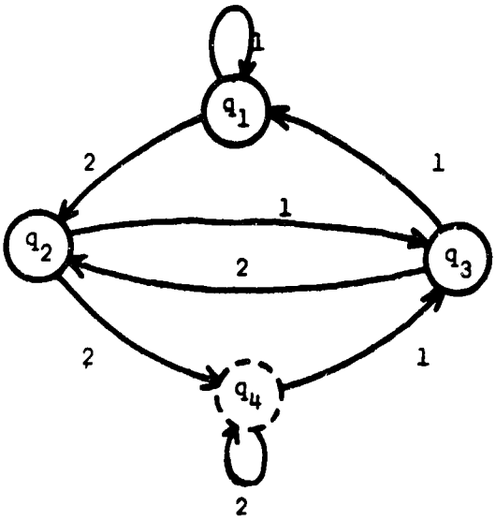
A.1 Introduction

This Appendix describes a method for the generation of Markov sequences by a small scale digital computer. The machine language computer program was written particularly for a Digital Equipment Corporation PDP-5 data processor, a 4096 12 bit word machine.

The order of the process, r, and the number of symbols, m, are completely general, and only limited by the available memory of the computer. The basic machine language program uses about 70 locations of core memory. A maximum of an additional $2m^r$ locations are required to store statistical information about the process being generated. This information must be stored in the computer memory prior to execution of the program. One step of this involves converting probabilities to coded numbers which are used by the computer.

A.2 Theory of operation

An r-th order Markov process - one whose present output depends on at most the past r outputs - may be described by a state diagram containing $n=m^r$ states, where m is the number of different output symbols allowed. The states correspond to all possible r-length sequences of the m output symbols. For each state m conditional probabilities must be specified to define the "next state" transitions of the process. An example of such a state diagram is given in figure A.1 for a second-order process (r=2) with 2 possible output symbols (m=2). The conditional probabilities are derived from those listed in Table A.1. It should be noted that certain states (shown in dotted lines in the diagram) have no transitions into them; they are never reached, and may be eliminated from



11 → q_1

12 → q_2

21 → q_3

22 → q_4

Figure A.1

Typical Transition Diagram of A Markov Process

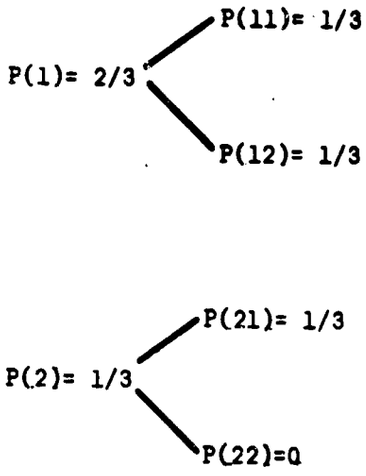


Table A.1

Probability Tree Associated

With A Markov Process

the diagram. Hence,

$$n \leq m^r.$$

Also, not every state has m transitions from it. This may happen in certain processes when the transition probabilities for these cases are zero. Nevertheless the sum of all transition probabilities from any state is always unity. In the computer program, which orders the transitions from lowest to highest probability, some states must be specified for these non-existent transitions and assigned a zero probability; the actual states specified are of no importance since the transitions will never occur.

To generate a Markov process the computer needs all the information contained in the state diagram. This is:

- number of states, n.
- number of output symbols, m.
- ordered listing of next states and corresponding probabilities for every state of the process.
- coded numbers corresponding to the transition probabilities.
- starting state.
- outputs corresponding to each state.

The flow diagram of Figure A.2 describes the operation of the program in the generation of Markov sequences.

The "next state" transitions of the process are determined by sampling a Gaussian noise generator connected by an analog-to-digital convertor to the computer, adding to the sample a constant which corresponds to the probability of going to the least likely state, and checking the magnitude of the resultant binary number to see if it is above or below a specified limit. If the limit is exceeded, the particular state corresponding to the constant added to the sample is specified as

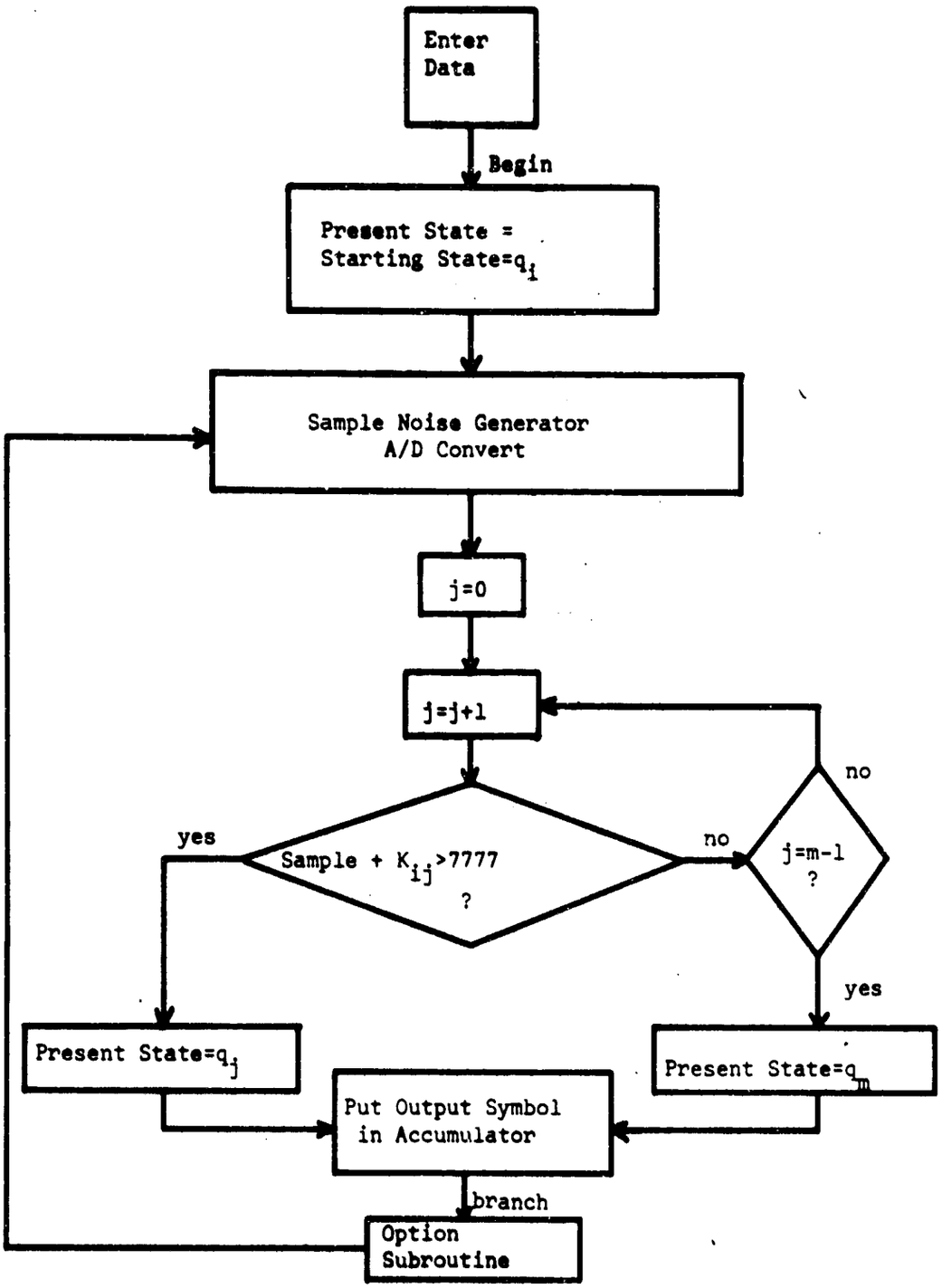
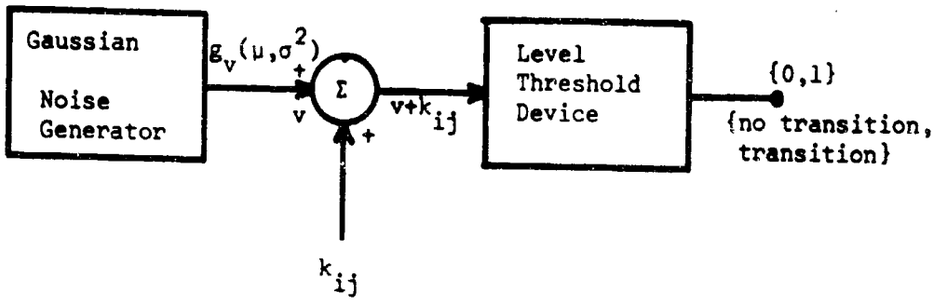


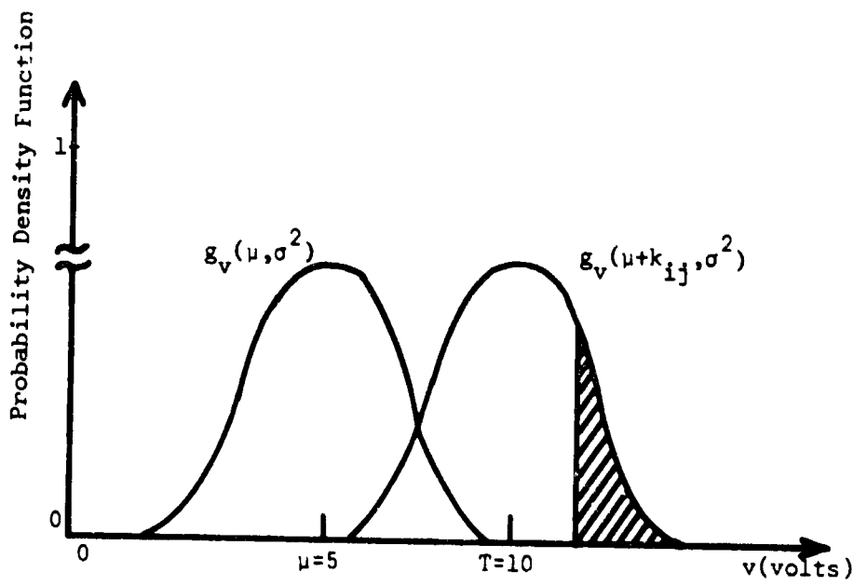
Figure A.2

Flow Diagram-Markov Sequence Generator Program

the next state of the process. If the limit is not exceeded, a second constant (corresponding to the next most likely state transition) is added to the same sample and again checked in a like manner. This procedure continues until the limit is exceeded; the state corresponding to the added constant is taken as the next state. Only m-1 iterations at most are necessary to effect a state transition, since, if m-1 states are not chosen as the next state, the m-th ordered state must be. The state transitions are always checked from least probable to most probable, thus the necessity for their entry in an ordered manner. The constants which are added to the a/d converted sample of the noise generator are those which are stored in memory prior to execution, and correspond to shifting the mean of the Gaussian noise source to a point where the desired transition probabilities are obtained by the given decision rule.

Figure A.3 demonstrates how the statistical properties of the noise source are related to the transition probabilities of the Markov process. The noise source has a mean μ of 5 volts and a standard deviation σ of 1 (variance $=\sigma^2=1$). The binary conversion of any sample between 0 and 10 volts corresponds to the octal numbers 0000 through 7777. The computer program checks to see if the constant K_{ij}, corresponding to P_{ij} (the probability of a transition from state i to the state corresponding to the j-th ordered probability) plus the noise sample exceeds 7777. This procedure is equivalent, in the analog case, to seeing if a voltage k_{ij} (the analog equivalent of the binary constant, K_{ij}, actually used in the program) plus the noise sample voltage, v, produces a result greater than 10 volts. We denote the Gaussian density function of the random variable v corresponding to a distribution with a mean of μ and a





$$\int_T^\infty g_v(\mu+k_{ij},\sigma^2)dv = P(\text{Transition State-i} \rightarrow \text{State-j})$$

Figure A.3

Noise Source and Transition Probabilities of Markov Source

variance of σ^2 by $g_v(\mu;\sigma^2)$, we see (//) that the following probabilities correspond to the sample plus constant, $v+k_{ij}$, being greater or less than 10 volts:

$$1-P_{i,j}=P[v+k_{i,j}<10v]=\int_{-\infty}^{10} g_v(5+k_{i,j};1)dv=\frac{1}{2}+\text{erf}[5-k_{i,j}] \quad \text{A.2.1}$$

$$P_{i,j}=P[v+k_{i,j}\geq 10v.]=\int_{10}^{\infty} g_v(5+k_{i,j};1)\,dv=\frac{1}{2}-\text{erf}[5-k_{i,j}] \quad \text{A.2.2}$$

Equation A.1 is just the probability of not choosing the transition corresponding to $k_{i,j}$, while equation A.2 is the probability of choosing it. If this particular transition is not choosen, it is necessary to see if the transition with the next highest probability will cause the sample plus constant to be greater than 7777 (octal), i.e., $v+k_{i,j+1}\geq 10$ volts. We must remember, however, that we know from the j-th iteration that $v<10-k_{i,j}$ and, so, the constant which is added must be sufficient so that:

$$P[10\leq v+k_{i,j+1}<10-k_{i,j}+k_{i,j+1}]=P_{i,j+1}$$

But, $P[10\leq v+k_{i,j+1}<10-k_{i,j}+k_{i,j+1}]=\int_{10}^{10-k_{i,j}+k_{i,j+1}} g_v(5+k_{i,j+1};1)dv$

$$=\text{erf}[(10-k_{i,j}+k_{i,j+1})-(5+k_{i,j+1})]-\text{erf}[10-(5+k_{i,j+1})]$$

$$=\text{erf}[5-k_{i,j}]-\text{erf}[5-k_{i,j+1}]$$

and from the j-th iteration we know:

$$P_{i,j}=\frac{1}{2}-\text{erf}[5-k_{i,j}];\ \text{or erf}\,[5-k_{i,j}]=\frac{1}{2}-P_{i,j}$$

Thus, $P_{i,j+1}=[\frac{1}{2}-P_{i,j}]-\text{erf}\,[5-k_{i,j+1}]$

or, $P_{i,j}+P_{i,j+1}=\frac{1}{2}-\text{erf}[5-k_{i,j+1}]=\int_{10}^{\infty} g_v(5+k_{i,j+1};1)dv$ A.2.3

We must, then, choose $k_{i,j+1}$ so that equation A.3 is satisfied; the procedure is to add $P_{i,j}$ to $P_{i,j+1}$, subtract this sum from 1/2, and use tables of the Gaussian error function to find $k_{i,j+1}$. If the transition to the

j+1 ordered state is not made, the next iteration will use $k_{i,j+2}$ added to the same sample. and by the above reasoning we may find $k_{i,j+2}$ by the equation: $k_{i,j+2} = 5\text{-erf}^{-1}[\frac{1}{2} - (P_{i,j} + P_{i,j+1} + P_{i,j+2})]$ volts A.2.4

In general: $k_{k,j+w} = 5 - \text{erf}^{-1}-[(\sum_{\xi=0}^{w} P_{i,j+\xi}) - \frac{1}{2}]$ volts A.2.5

The actual constant used by the computer is the a/d converted binary number corresponding to this voltage. Note that $\text{erf}^{-1}(x)$ may run from $-\infty$ to $+\infty$ for values of x equal to $-\frac{1}{2}$ and $+\frac{1}{2}$ respectively. The a/d converter, however, is limited to a 10 volt range, and this restriction must be imposed on the voltage $k_{i,j+w}$. This approximation causes no problem, however, since erf 3.87= 1/2 when rounded off beyond 4 places, and this corresponds to the limits of only 1.13v and 8.87 v respectively. The binary numbers 0000 and 7777 (correspond to 0 and 10 volts) may be used for the probability of zero and one respectively.

Once a transition is made, the same process is repeated but uses the set of probabilities and state transitions which were entered for that particular state. The process continues to generate next state transitions with the desired probabilities until the program is halted by the operator or control is removed by programming in a special subroutine described below.

After each state transition, the main program branches to a subroutine (written by each user) which allows the present state information to be used in producing the desired output information in the required form. Some possible options might be:

- store a sequence of outputs for future processing by another program.
- convert the output information to an analog voltage, and hold this voltage on an output line.

- activate particular relays or control circuits which correspond to the various states of the Markov process.

Also by proper programming within this subroutine, control may be removed from the Markov Sequence Generator Program and transferred to some other location.

Since output assignments occurs after state transitions occur, and the state-to-output mapping may be specified in any way, it is possible for the output process to be a projection of a Markov process, or, in general a Linearly Dependent Process. A discussion of the properties of such statistical processes is presented in more detail by Booth (9).

Appendix B

Miscellaneous Derivations

B.1 Steady State Probabilities of A Markov Source

In this section the steady-state probabilities associated with a Markov process will be found and a useful form presented for a first-order binary Markov process.

Let T be the transition matrix of a Markov process.

$$T = [t_{ij}] \qquad \text{B.1.1}$$

Each element, t_{ij}, represents the probability of a transition from state-i to state-j, where the states may be assumed to correspond to the past r output symbols for an r^{th}-order process. Figure B.1 shows the transition diagram for a first-order binary Markov process.

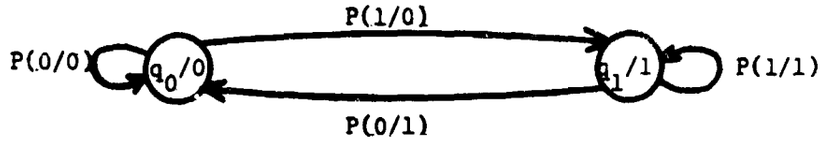


Figure B.1

Following the presentation by Booth (9), let $\pi_i(n)$ be the probability that the system is in the i^{th} state at the n^{th} observation. The probability (row) vector $\bar{\pi}(n)$ represents the probability of the system being in each state at observation n. The probability vector at observation n+1 is related to the probability vector at observation n by the matrix equation,

$$\bar{\pi}(n+1) = \bar{\pi}(n)\, T \qquad \text{B.1.2}$$

It is assumed here that the elements of T are time invariant. Thus,

$$\begin{aligned} \bar{\pi}(n+1) &= \bar{\pi}(n)\, T = [\bar{\pi}(n-1)\, T]\, T \\ &= \bar{\pi}(n-1)\, T^2 \end{aligned} \qquad \text{B.1.3}$$

or in general,

$$\bar{\pi}(n+1) = \bar{\pi}(0)\, T^{n+1} \qquad \text{B.1.4}$$

or

$$\bar{\pi}(n) = \bar{\pi}(0)\, T^{n} \qquad \text{B.1.5}$$

It is shown by Booth (9) that the z-transform of T^n is,

$$Z[T^n] = W(z) = z\,[zI - T]^{-1} \qquad \text{B.1.6}$$

where I is the identity matrix.

Using the final value theorem of z-transforms, we may write,

$$\lim_{n\to\infty} [T^n] = \lim_{z\to 1} (z-1)\, W(z) = \lim_{z\to 1} (z-1)(z)[zI-T]^{-1} \qquad \text{B.1.7}$$

Consider, now, specifically the following binary first order Markov process:

$$T = \begin{vmatrix} a & 1-a \\ 1-b & b \end{vmatrix} \qquad \text{B.1.8}$$

$$[zI - T] = \begin{vmatrix} z-a & a-1 \\ b-1 & z-b \end{vmatrix} \qquad \text{B.1.9}$$

$$[zI - T]^{-1} = \frac{1}{z-1} \begin{vmatrix} \frac{z-b}{z-c} & \frac{z-a}{z-c} \\ \frac{1-b}{z-c} & \frac{z-a}{z-c} \end{vmatrix} \qquad \text{B.1.10}$$

where c = a+b-1. And from the final value theorem,

$$\lim_{z\to 1} (z-1)(z)[zI-T]^{-1} = \begin{vmatrix} \frac{1-b}{1-c} & \frac{1-a}{1-c} \\ \frac{1-b}{1-c} & \frac{1-a}{1-c} \end{vmatrix} = \lim_{n\to\infty}[T^n] \qquad \text{B.1.11}$$

Notice that this matrix has identical rows, and hence the steady-state probability vector is,

$$\lim_{n\to\infty} \bar{\pi}(n) = \lim_{n\to\infty} \bar{\pi}(0)\, T^n = \left[\frac{1-b}{1-c} \quad \frac{1-a}{1-c} \right] \qquad \text{B.1.12}$$

B.2 Dependence of Weighting Factors in A First-Order Markov Optimum Detector

In section 3.4 the following expression was obtained for the optimum detector of patterns which may have been generated by either a statistically independent process or a first-order Markov process:

$$N\log P(0/0) + N_1 \log \frac{P(1/0)P(0/1)}{P^2(0/0)} + N_{11} \log \frac{P(0/0)P(1/1)}{P(0/1)P(1/0)} \qquad \text{B.2.1}$$

$$\geq (t^{-1})\log 1/2 : D_1$$

$$\text{otherwise} : D_2$$

Let $C_1 = \frac{P(1/0)P(0/1)}{P^2(0/0)}$ and $C_{11} = \frac{P(0/0)P(1/1)}{P(0/1)P(1/0)}$

It is shown in this section that for a fixed first-order probability distribution, P(0) and P(1), C_1 and C_{11} may not vary independently. The relation between these factors, and thus the form of the displays which may be generated by such processes is also indicated.

Let the transition matrix for a first-order Markov process be,

$$T = \begin{vmatrix} a & 1-a \\ 1-b & b \end{vmatrix} \qquad \text{B.2.2}$$

where the t_{ij} entry represents the probability of a transition from state-i to state-j. If the states associated with this matrix are chosen to correspond with the output symbols of the process, it is possible to write C_1 and C_{11} as,

$$C_1 = \frac{(1-a)(1-b)}{a^2} \quad \text{and} \quad C_{11} = \frac{a \cdot b}{(1-a)(1-b)} \qquad \text{B.2.3}$$

Let the product of C_1 and C_{11} be Q,

$$Q = C_1 \cdot C_{11} = b/a \qquad \text{B.2.4}$$

Assume, now, that C_1 may be held constant while C_{11} is allowed to vary. The product, Q, should be able to take on two possible values, Q_1 and Q_2, corresponding to the two values C_{11}^1 and C_{11}^2.

$$Q_1 = C_1 \cdot C_{11}^1 = b_1/a_1$$
$$Q_2 = C_1 \cdot C_{11}^2 = b_2/a_2 \qquad \text{B.2.5}$$

However, in section B.1 it was shown that the steady state probability vector is,

$$\bar{\pi}(\infty) = [P(0) \quad P(1)] = \left[\frac{1-b}{1-c} \quad \frac{1-a}{1-c}\right] \qquad \text{B.2.6}$$
$$= \left[\frac{1-b}{(1-a)+(1-b)} \quad \frac{1-a}{(1-a)+(1-b)}\right]$$

Thus, for a fixed P(0) and P(1), their ratio is,

$$P(0)/P(1) = (1-b) / (1-a)$$

and is constant. So, b/a is also constant. But this contradicts the assumption that C_{11} may vary independently of C_1.

It has been shown in this section that C_1 and C_{11} may not vary independently. In fact, once the steady-state (first order) probabilities are set, the ratio b/a is set, which determines the relation between C_1 and C_{11}. Note, however, that b and a may vary over wide ranges for a constant b/a ratio. It is necessary to insure only that

$$0 \leq (a \ \& \ b) \leq 1$$

B.3 Value of C_1 and C_{11} for P(0) = P(1) = 1/2

When it is desired that $\bar{\pi}(\infty) = [1/2 \quad 1/2]$, what values may C_1 and C_{11} take on? From relation B.1.12 the condition P(0)=P(1) implies

$$1-b = 1-a \text{ , or } \quad a=b$$

Thus the transition matrix becomes,

$$T = \begin{vmatrix} a & 1-a \\ 1-a & a \end{vmatrix} \qquad \text{B.3.1}$$

This results in,

$$C_1 = \frac{(1-a)^2}{a^2} \quad \text{and} \quad C_{11} = -\frac{a^2}{(1-a)^2} \qquad \text{B.3.2}$$

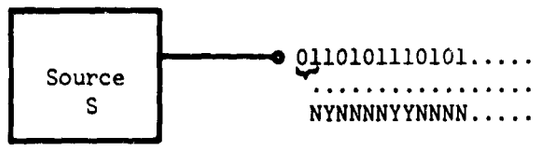
$$\text{or} \quad \log C_1 = -\log C_{11} \qquad \text{B.3.3}$$

B.4 Distribution of Second-Order Sequence Counts

Let P(0) and P(1) be the first-order probabilities of a 0 and a 1 respectively, and P(0/0), P(0/1), P(1/0), and P(1/1) be the conditional probabilities associated with the output symbols. Assume that dependencies extend only to the adjacent symbols. This describes the statistics of a first-order binary Markov process. The problem is to determine the distribution of N_{11}, the number of 11 sequences which occur in a longer sequence of length N.

If one observes the symbols generated by the Markov process one at a time, the chance of a symbol being a 1 is just P(1). The distribution of N_1, the number of 1's in an N length sequence, is binomial with mean of N·P(1). Now, consider the symbols emitted by the Markov process two at a time as depicted in Figure B.2. Each pair of symbols may be classified as being a 11 sequence (Y) or not being a 11 sequence (N).

Figure B.2



The problem has been transformed into determining the distribution of the Y's in the classified sequence. A Y occurs only when a 11 occurs, so P(Y) = P(11) = P(1)P(1/1). But the Y's, and hence the 11's, are obviously binomially distributed with mean of N·P(11). The

same approach may be used to find the distribution of any other sequences of length greater than or equal to one.

B.5 Entropy of a Markov Source

Consider a first-order Markov information source, S, which has an output symbol set $s_1, s_2, \ldots s_m$, with associated symbol probabilities $P(s_1), P(s_2), \ldots P(s_m)$, and the set of conditional symbol probabilities $\{P(s_i/s_j), i,j = 1,m\}$. The entropy (10) of a Markov source is defined as,

$$H(S) = \sum_{i,j=1}^{m\ m} P(s_i s_j) \log_2 1/P(s_i/s_j) \qquad \text{B.5.1}$$

Rewriting the conditional probabilities,

$$H(S) = -\sum_{i,j=1}^{m\ m} P(s_i s_j) \log P(s_i s_j)/P(s_j) \qquad \text{B.5.2}$$

which may be split into two terms,

$$H(S) = -\left[\sum_{i,j=1}^{m\ m} P(s_i s_j) \log P(s_i s_j) - \sum_{i,j=1}^{m\ m} P(s_i s_j) \log P(s_j) \right] \qquad \text{B.5.3}$$

The second summation may immediately be taken over i, giving,

$$H(S) = -\sum_{i,j=1}^{m\ m} P(s_i s_j) \log P(s_i s_j) + \sum_{j=1}^{m} P(s_j) \log P(s_j) \qquad \text{B.5.4}$$

Let S^2 be a source which has m^2 output symbols, σ_i, composed of pairs of output symbols of S, with symbol probabilities of,

$$\begin{aligned} P(\sigma_1) &= P(s_1 s_1) \\ P(\sigma_2) &= P(s_1 s_2) \\ P(\sigma_3) &= P(s_1 s_3) \\ &\vdots \\ P(\sigma_{m^2}) &= P(s_m s_m) \end{aligned} \qquad \text{B.5.5}$$

Furthermore, call $\overline{S}$ the adjoint of S, and let it be a source which has identical first order probabilities as S, but no dependencies, i.e., a statistically independent process. Let $\overline{S^2}$ be the adjoint of S^2. Equation B.5.4 may be written in terms of these special sources as,

$$H(S) = H(\overline{S^2}) - H(\overline{S}) \qquad \text{B.5.6}$$

Bibliography

1. Kaufman, H.M., R.M. Levy, T.L. Booth, R.M. Glorioso, "An Experimental Investigation of An Integrated Man-Computer Signal Detection System," Presented at Annual Meeting of IEEE Human Factors in Electronics, May 1965.

2. Brazeal, E.H., Jr., and T.L. Booth, "Operator Noise in a Discrete Signal Detection Task," IEEE Transactions on Human Factors in Electronics, Vol. HFE-7, No. 4, Dec. 1966, pp164-173.

3. Moran, J., and T.L. Booth, "Operator Performance in A Forced Choice Visual Detection Experiment," M.S. Thesis, University of Connecticut, 1967.

4. Glorioso, R.M., "A Stochastic Model of the Human Observer As A Detector of Signals Embedded In Noise," Ph.D. Dissertation, University of Connecticut, 1967.

5. Julesz, B., "Visual Pattern Discrimination," IRE Transactions on Information Theory, February, 1962.

6. Booth, T.L., R.M. Glorioso, R.M. Levy, J.R. Walter, and H.M. Kaufman, "Displays for Studying Signal Detection and Pattern Recognition," Presented at Spring DECUS (Digital Equipment Computer Users Society) Meeting, 1967.

7. Brazeal, E.H., and T.L. Booth, "A High-Speed Man-Computer Communication System," Presented at Spring DECUS (Digital Equipment Computer Users Society) Meeting, 1965.

8. Selin, I., Detection Theory, Princeton University Press, 1965.

9. Booth, T.L., Sequential Machines and Automata Theory, J. Wiley and Sons, 1967.

10. Abramson, N., Information Theory and Coding, McGraw- Hill, 1963.

11. Papoulis, A., Probability, Random Variables, and Stochastic Processes, McGraw-Hill, 1965.

GD/EB-25

UNIVERSITY OF CONNECTICUT

STORRS, CONNECTICUT

GENERAL DYNAMICS/ELECTRIC BOAT RESEARCH PROJECT

PREDICTION OF REACTION TIME FROM INFORMATION OF INDIVIDUAL STIMULI

by

Jerry Lamb

Electric Boat division

Herbert M. Kaufman

Department of Psychology

July 1967

Acknowledgement is due the Office of Naval Research which in part supported this program through a prime contract (NOnr 2512(00)) with Electric Boat division of General Dynamics as a part of the SUBIC (Submarine Integrated Control) program.

PREDICTION OF REACTION TIME FROM INFORMATION OF INDIVIDUAL STIMULI

1.0 Introduction

There has been continuing disagreement in the literature over the effects on reaction time of the information load of unequally probable stimuli. It has been shown that while RT is linear with average stimulus information, the same function does not apply with regard to the information of individual stimuli. Kaufman and Lamb (1966) advanced the hypothesis that S's behavior in this type of situation is a function of his threshold for differential stimulus probabilities. Their experiment differed from previous studies on two variables. First, they used only two stimuli for all conditions in which stimuli were not equally probable; and second, they used an absolute judgment situation, where other studies have used discriminative judgments. The present study was conducted to explore the significance of the number of equally probable and unequally probable stimuli, to test the validity of Kaufman and Lamb's hypothesis, and to attempt to modify the hypothesis to allow quantitative predictions. The experiment varied the number of unequally probable stimuli in a discrimination setting and was designed to follow as closely as possible the procedure used by Hyman (1953).

2.0 Experiment

The Ss were 48 male and female undergraduates. The apparatus consisted of a Gebrand tachistoscope, voice key, and Hunter millisecond timer. Stimuli consisted of white cards with black stimuli, X's and O's, 7/8 inch in size.

The stimulus locations used were the four outermost corners and next inner four corners of an imaginary 6x6 matrix. Bun, boo, bee. bore, bive, bix, bev, and bate were the eight location names of which two, four, or all were used, depending on the condition. Each side of the matrix made a visual angle of approximately 5^{o} at S's location. The matrix was centered on the white card.

The data for the information in the individual stimuli are of the same form as that reported by Hyman, that is, RT to high probability stimuli are longer than would be predicted from the regression line for equally probable stimuli, and the reverse for low probability stimuli. Figure 1 shows that stimuli with the same probability of occurrence (7/8 or 1/2) had approximately the same RT regardless of the number of alternatives in the condition.

While the data are of the same general form as that reported by Hyman (1953) and Kaufman and Lamb (1966), the present results provide quantitative values for testing an extension of the hypothesis advanced by Kaufman and Lamb. They had proposed that, in an absolute judgmen situation with unequal probabilities, Ss would

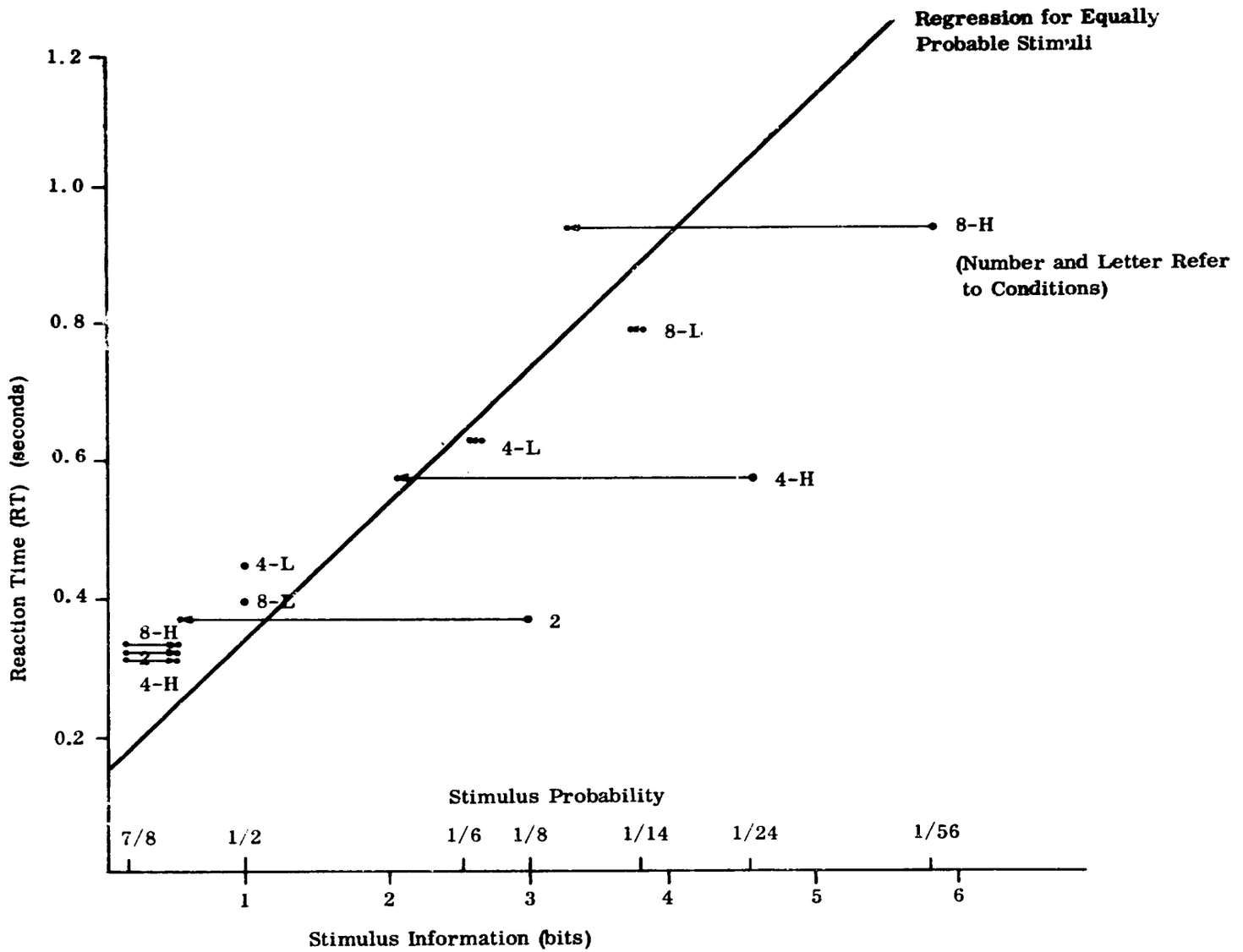


Figure 1. RT to Stimulus Probability for Unequally Probable Conditions

be prepared to respond with the name of the more frequent stimulus provided that the disparity in probabilities was large enough and the cost of a mistake was not excessive. The extension is that, with three or more alternatives, S makes a chain of decisions, the order of which depends on the probabilities of the alternatives and the time for each of which depends on the amount of information in each step. On every trial, S makes an initial decision as to whether the most probable stimulus has occurred. If it has, then S's reduction in uncertainty is equivalent to the information in the most probable stimulus plus the residual information in all remaining stimuli. Thus, for a set of stimuli, 1 to n, ranked in order of probability, the reduction in probability for the most probable stimulus is

$$-p_1 \log_2 p_1 - (1-p_1) \log_2 (1-p_1) \qquad (1)$$

Note that the second term (residual information) is not the same as average information.

If the most probable stimulus does not occur, then the time required for this decision is the time that S uses to process the information in the first term of eq. 1. Next, S decides if the second most probable stimulus has occurred. The total reduction for the second most probable stimulus occurring is

$$-p_1 \log_2 p_1 \; -p_2 \log_2 p_2 - (1-p_1-p_2) \log_2 (1-p_1-p_2)$$

or first stimulus reduction plus second stimulus reduction plus residual information. This process is repeated until a decision has been made for all stimuli.

If, at any point, the remaining stimuli are all equally probable, the residual term is simply $\log_2$ of the number of stimuli remaining. Thus, for the present experiment, two equations are sufficient, eq. 1 for the most probable stimulus and

$$-p_1 \log_2 p_1 + \log_2 n \text{ remaining} \qquad (2)$$

for all other stimuli.

The reduction in information for each stimulus was calculated and the RT to that amount of information was estimated from the regression line for equally probable alternatives. Table I gives predicted and actual RTs for the present experiment; t - tests were used to test for significant departures from predicted values. None were found to be significant. Table I also shows values estimated from other published data; these values are consistent with the results obtained in this study.

Thus, for discrimination situations at least, a quantitative method using only the information loadings of individual stimuli can predict RTs to individual unequally probable stimuli.

An experiment has been conducted using the same conditions for absolutely judged stimuli. Preliminary results are of the same form as the present experiment.

Table I. Reduction in Uncertainty, Predicted and Actual RTs for Three Studies

	EXPERIMENTAL CONDITIONS			
CONDITION	NO. ALTER-NATIVES	PROB-ABILITY	STIMULUS INFORMATION	AVERAGE INFORMATION
1) 2ELA	2	1/2 (.500)	1.0	1.0
2) 4ELA	4	1/4 (.250)	2.0	2.0
3) 8ELA	8	1/8 (.125)	3.0	3.0
4) 2ULA	1	7/8 (.875)	0.1926	0.5436
	1	1/8 (.125)	3.0	
5) 4ULA-High	1	7/8 (.875)	0.1926	0.7417
	3	1/24 (.042)	4.585	
6) 8ULA-High	1	7/8 (.875)	0.1926	0.8945
	7	1/56 (.018)	5.8074	
7) 4ULA-Low	1	1/2 (.500)	1.0	1.7925
	3	1/6 (.167)	2.585	
8) 8ULA-Low	1	1/2 (.500)	1.0	2.4037
	7	1/14 (.071)	3.8074	

PROBABILITY	N	TOTAL N	REDUCTION IN UNCERTAINTY	PREDICTED	ACTUAL
			Present Study		
.875	1	2	0.5436	.332	.315
.875	1	4	0.5436	.332	.314
.875	1	8	0.5436	.332	.324
.500	1	4	1.0	.335	.440
.500	1	8	1.0	.335	.397
.167	3	4	2.585	.654	.630
.125	1	2	0.5436	.332	.370
.071	7	8	3.807	.900	.786
.042	3	4	2.129	.562	.572
.018	7	8	3.351	.808	.937

Table I. (Cont'd)

PROBABILITY	N	TOTAL N	REDUCTION IN UNCERTAINTY	PREDICTED	ACTUAL
			Hyman (1953)		
.813	1	4	0.69	.317	.306
.062	3	4	1.72	.475	.585
			Stone & Calloway (1964)		
.812	1	4	0.695	.326	.325
.188	3	4	1.827	.370	.375
.500	1	4	1.0	.338	.345
.250	1	4	1.5	.357	.370
.125	2	4	2.0	.377	.375

GB/EB 26

UNIVERSITY OF CONNECTICUT

STORRS, CONNECTICUT

GENERAL DYNAMICS/ELECTRIC BOAT RESEARCH PROJECT

ALPHABET SIZE, IMPLICIT CODING AND THE MEMORY SPAN

by

Robert M. Levy
Electric Boat division

Herbert M. Kaufman
Grace Vogelli
Department of Psychology

September 1968

Acknowledgement is due the Office of Naval Research which in part supported this research through a prime contract (NOnr 2512(00)) with General Dynamics/ Electric Boat as a part of the SUBIC (Submarine Integrated Control) program.

ALPHABET SIZE, IMPLICIT CODING AND THE MEMORY SPAN

1.0 Introduction

Miller (1956) suggested that the immediate memory span (IMS) is constant for chunks, where a chunk represents a unit of response. Hyman and Kaufman's (1966) results indicated, to the contrary, a constancy for information. The present experiment was an attempt to resolve the disagreement between the two sets of results. Miller's interpretation is based primarily on the IMS for sequences of familiar binary stimuli presented aurally. With the use of various coding schemes, the IMS increased for the amount of information transmitted but was constant for the number of response units or chunks. Hyman and Kaufman presented tachistoscopically simultaneous messages of 4 to 8 symbols selected from alphabets of either 3 or 5 bits per symbol. Their symbols were either eight forms (3-bit alphabet) or combinations of the forms with four colors (5-bit alphabet), and the messages were exposed for either 100 msec. or 500 msec. No significant differences were found in the number of bits recalled, approximately 13.3, as functions of the exposure time or bits per symbol conditions.

Hyman and Kaufman (1966) suggested that the difference between the two sets of data might be in the human's ability to code stimuli. With familiar stimuli, such as in Miller's experiments, Ss might be able to encode them during the brief interval that an exposure remains. The typical sequential presentation allows relatively large amounts of time for coding.

Two parameters appear to be of fundamental interest, viz., alphabet size and familiarity. The chunk hypothesis is based on binary alphabets of familiar symbols. Sperling (1960) found a constant IMS of 4.5 symbols for brief tachistoscopic exposures of messages selected from alphabets of either 21 consonants or 21 consonants plus 10 digits. Hyman and Kaufman's results are based on relatively unfamiliar alphabets of 3 and 5 bits per symbol. Therefore, in order to resolve this contradition, certain features of the Hyman and Kaufman experiment were replicated with familiar symbols and an alphabet of size two was included. In the present experiment, alphabet size was varied from 1 to 4.7 bits per symbol with familiar symbols — letters of the English alphabet.

2.0 Method

2.1 Stimuli, Apparatus, and Subjects

Familiar subsets of letters from the English alphabet were selected to give "alphabets" of 2, 4, 8, 16, and 26 alternatives corresponding to 1, 2, 3, 4, and 4.7 bits of information per symbol (table I). The sets were the letters A-B, A-D, A-H, A-P, and A-Z. Messages were always of length 12 and were formed by random sampling with replacement. Fifty messages were prepared for each alphabet. For the two-alternative case, the distribution of number of symbols on each card followed

Table I. Alphabet Sets

NUMBER OF SYMBOLS	BITS/SYMBOL	LETTER SET
2	1	A-B
4	2	A-D
8	3	A-H
16	4	A-P
26	4.7	A-Z

a binomial distribution. The letters were printed on 8-1/2 x 11 inch white cards using a primer print typewriter. Each letter was 1/4 inch high and 1/8 inch across. The 12 symbols were arranged in a diamond 2 x 2-1/8 inches, which subtended an angle of approximately 5° (figure 1). The cards were presented in a Gebrands two-field tachistoscope. The second field contained a center fixation point and was brightly lit to minimize afterimages.

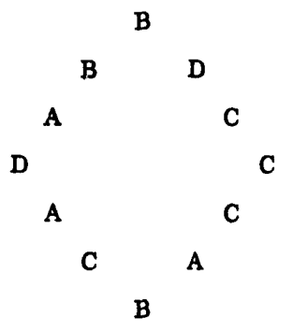


Figure 1. Symbol Arrangement

Two groups of Ss were run. In the first group of 10 Ss, 2 Ss were assigned randomly to each of the five alphabet conditions. In a single session, each S saw 100 messages in a single condition. In the second group, each of 5 Ss observed all of the conditions four times over the period of ten sessions.

2.2 Procedure

The S was seated at the tachistoscope in a darkened room and asked to fixate on the fixation point. He then initiated a trial by pushing a button which exposed the stimulus for 500 msec. The S was then given as much time as he needed to write down the symbols on a response grid. The Ss were instructed to not guess.

3.0 Results and Discussion

For the 2-, 4-, 8-, 16-, and 26-symbol alphabets, the average number of symbols recalled were (figure 2):

	Alphabet Size				
	2	4	8	16	26
Repeated measures	5.2	4.5	4.1	4.2	4.1
Independent groups	5.1	3.7	4.1	3.7	4.0

The most striking feature is the nearly constant level for all other conditions following a decrease from the level for the two-symbol alphabet. Since there were no

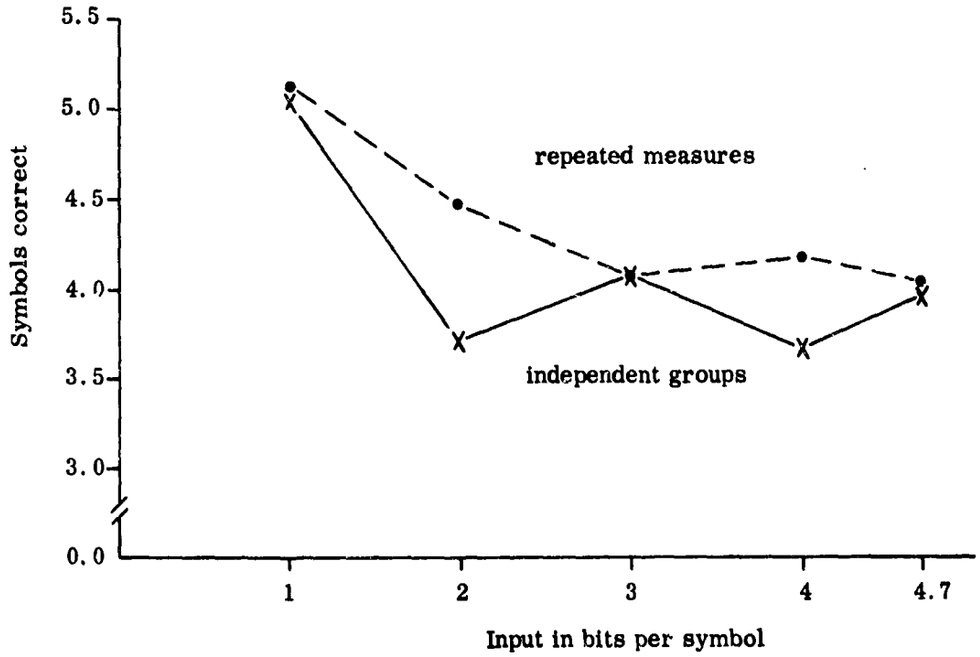


Figure 2. Symbols Recalled as a Function of Alphabet Size

coding conditions, the number of symbols recalled corresponds to the number of chunks recalled. With the exception of the binary alphabet, the data support the chunk constancy hypothesis. Certainly, a constancy for information is out of the question. Even the deviant (alphabet size of two) data may be explained within the framework of the chunk concept. The explanation is based on the assumption that the chunk is not the single symbol in the binary case. With a binary alphabet, the S is able to increase his IMS by invoking a simple implicit coding procedure.

One possible code is to operate on the basis of runs, i.e., sequences of the same symbol. A coded response would be 2A, 4B, 1A, etc. This code has two

response units per chunk, one specifying number and one specifying kind. Therefore, the S would use it only when the run length is greater than two. Analyzing the distributions of samples of size 12 for the binary alphabet, we find that the probability of a run of length two or less is .79, incorporating .56 of the symbols. Calculating from the data (repeated measures) for the larger alphabets of 8, 16, and 26 symbols, we find that the number of response units available is 4.1. The number of recalled symbols used by runs of length two or less is .56 x 4.1 = 2.3. The runs of length greater than two, which average 3.6, are divided into the remaining 1.8 response units giving 3.2 symbols for the 1.8 response units. Adding 3.2 to 2.3, we obtain 5.5 as the predicted number of symbols to be recalled in the two-symbol alphabet condition. The observed figure of 5.2 is close enough to the predicted figure to support the notion that some such process might be operating. The hypothesized implicit coding strategy is most useful with two alternatives. However, some gain would be expected for a four-symbol alphabet. Thus, Ss run repeatedly show a slightly better performance for the four-alphabet condition; this may be a systematic effect enhanced by practice. The explanation of some forms of information processing behavior in terms of repetition has been proposed previously by Kornblum (1967).

Thus, the results of the present experiment support Miller's hypothesis that the IMS is constant for "chunks" or units of response. If this is the case, Hyman and Kaufman's (1966) data are open to reinterpretation. They found that the maximum number of symbols correctly recalled was 4.5 in the 3-bit per symbol form group. Two points are important about this: First, the number of symbols recalled is the same as found in the present experiment for comparable conditions and also found by Sperling (1960). Secondly, the stimulus figures were complex forms which may well have been as distinctive and, with some training, as familiar as the letters used in the present experiment. The maximum number of symbols recalled for the 5-bit per symbol color-form conditions was approximately 2.75, well below the comparable figure for the present experiment. Inspection of Hyman and Kaufman's data suggests that, at least for the 500 msec. exposure condition, asymptotic performance was not obtained. Whereas the form-alone alphabet was relatively familiar, it may be that the color-form alphabet was relatively novel. Unfortunately, the two alphabets are not comparable. If performance is still improving at the end of their experimental sessions in the color-form condition, then Hyman and Kaufman's interpretation of their data is open to doubt. However, one feature of their data may support the contention of a constancy for bits. For the 100-msec. exposure conditions, which were run after the 500-msec. conditions and were, therefore, more practiced, there is no evidence of a further increase in IMS over the final two sessions. If it should be the case that the interpretation applied to the 500-msec. exposure group is correct but that their 100-msec. exposure group had, in fact, stopped improving, then the significance of the exposure times becomes crucial. Possibly the 500-msec. exposure is already allowing processes to be invoked different from those available with the 100-msec. exposure.

Alternatively, the chunking and information capacities represent different limits on the organism's IMS. The information capacity may not have been reached in the present experiment because of the familiarity of the symbols. Therefore, a chunk

capacity was imposed by some other process in memory. Finding larger alphabet sizes which provide homogeneous subsets poses a problem for further research.

We conclude that at least for the conditions tested, the results support a modified chunk hypothesis.

4.0 References

Hyman, L.M., & Kaufman, H.M. "Information and the Memory Span." Perception and Psychophysics, 1966, 1, 235-237.

Kornblum, S. "Choice Reaction Time for Repetitions and Nonrepetitions: A Re-examination of the Information Hypothesis. Acta Psychologica, 1967, 27, 178-187.

Miller, G.A. "The Magical Number Seven, Plus or Minus Two. Some Limits of Our capacity for Processing Information." Psychological Review, 1956, 63, 81-97.

Sperling, G. "The Information Available in Brief Visual Presentations." Psychological Monographs, 1960, 74 (11, Whole No. 498).

General Dynamics Corp. Electric Boat Division
Technical Report U417-68-098
EXPERIMENTAL INVESTIGATIONS OF MAN-MACHINE PROCESSING OF INFORMATION, VOLUME III

October 1, 1968, 154 Pages

The aim of this project is to provide basic knowledge of the methods which may be used by a man-computer system to detect the presence of a target, using data from a passive sonar receiver. This research consists of analytical studies to evaluate important system parameters and experimental investigations measuring operator performance under various operating conditions.

The first two reports in this volume describe the effects of pattern variations on human pattern recognition. The results measured the operator's ability to visually detect pattern differing in shape and to detect patterns generated by statistically dependent sequences.

The second two reports deal with basic human information processing and describe the testing of a predictive model for reaction time to visual stimuli and a test of the effects of number of stimuli on memory span.

General Dynamics Corp. Electric Boat Division
Technical Report U417-68-098
EXPERIMENTAL INVESTIGATIONS OF MAN-MACHINE PROCESSING OF INFORMATION, VOLUME III

October 1, 1968, 154 Pages

The aim of this project is to provide basic knowledge of the methods which may be used by a man-computer system to detect the presence of a target, using data from a passive sonar receiver. This research consists of analytical studies to evaluate important system parameters and experimental investigations measuring operator performance under various operating conditions.

The first two reports in this volume describe the effects of pattern variations on human pattern recognition. The results measured the operator's ability to visually detect pattern differing in shape and to detect patterns generated by statistically dependent sequences.

The second two reports deal with basic human information processing and describe the testing of a predictive model for reaction time to visual stimuli and a test of the effects of number of stimuli on memory span.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
GENERAL DYNAMICS CORPORATION Electric Boat division Groton, Connecticut	UNCLASSIFIED 2b. GROUP —

3. REPORT TITLE

EXPERIMENTAL INVESTIGATIONS OF MAN-MACHINE PROCESSING OF INFORMATION, VOLUME III

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Annual Report, July 1967 to July 1968

5. AUTHOR(S) (First name, middle initial, last name)

Taylor L. Booth et al

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 1, 1968	154	—

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
NOnr 2512 (00) b. PROJECT NO.	U417-68-098
c. d.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Department of the Navy Office of Naval Research Washington, D.C.

13. ABSTRACT

The aim of this project is to provide basic knowledge of the methods which may be used by a man-computer system to detect the presence of a target, using data from a passive sonar receiver. This research consists of analytical studies to evaluate important system parameters and experimental investigations measuring operator performance under various operating conditions.

The first two reports in this volume describe the effects of pattern variations on human pattern recognition. The results measured the operator's ability to visually detect patterns differing in shape and to detect patterns generated by statistically dependent sequences.

The second two reports deal with basic human information processing and describe the testing of a predictive model for reaction time to visual stimuli and a test of the effects of number of stimuli on memory span.

DD FORM 1 NOV 68 1473 (PAGE 1)

S/N 0101-807-6801

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT